

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: January 22, 2002, 16:23:40 ; Search time 40.57 Seconds  
(without alignments)  
1137.482 Million cell updates/sec

Title: US-09-485-529-7

Perfect score: 3194  
Sequence: 1 MKREYQDAGSGGGGGGMS.....TLGWHTRPLTATSAWRLAGP 623

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq\_1101.\*  
1: /SID8/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID8/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID8/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID8/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID8/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID8/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID8/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID8/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID8/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID8/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID8/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID8/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID8/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID8/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID8/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID8/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID8/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID8/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID8/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID8/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID8/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID8/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3194	100.0	623	20	AAV02540
2	2683	84.0	630	20	AAV02541
3	2189	68.5	425	20	AAV02539
4	1612	50.5	587	22	AAE02560
5	1612	50.5	587	22	AAE01892
6	1594	48.9	532	18	AAW30792
7	1594	48.9	533	21	AAW38575
8	1590	49.8	532	22	AAE02545
9	1590	49.8	532	22	AAE01907
10	1589	49.7	517	21	AAW38577
11	1589	49.7	518	21	AAW38576

12	1588.5	49.7	531	21	AAW28575	Arabidopsis SCla2.
13	1572	49.2	572	22	AAW31883	Amino acid sequenc
14	1568	49.1	572	22	AAW31884	Amino acid sequenc
15	1542	47.7	587	21	AAW28574	Arabidopsis SCla8.
16	883.5	27.7	256	20	AAV02538	Protein encoded by
17	812.5	25.4	277	18	AAW38193	Arabidopsis SCARC
18	797.5	25.0	262	18	AAW38194	Arabidopsis SCARC
19	673.5	21.1	138	20	AAV02544	Protein encoded by
20	670	21.0	668	21	AAW28595	Maize Scarecrow pr
21	600.5	18.8	653	18	AAW38178	Arabidopsis SCARC
22	600.5	18.8	653	21	AAW28569	Arabidopsis SCARC
23	599	18.5	428	19	AAW81753	Tomato Ls Protein.
24	589.5	18.5	482	18	AAW30795	Arabidopsis thalia
25	569.5	17.8	259	18	AAW30794	Arabidopsis thalia
26	554	17.3	384	21	AAW29710	Arabidopsis thalia
27	554	17.3	405	21	AAW29709	Arabidopsis thalia
28	554	17.3	413	21	AAW29708	Arabidopsis thalia
29	523.5	16.4	405	21	AAW48516	Arabidopsis thalia
30	513	16.1	313	21	AAW28577	Maize SC1m1. Zea
31	512	16.0	313	18	AAW38209	Maize Zm-SC12 prot
32	512	16.0	313	21	AAW28573	Maize SC1m1. Zea
33	507	15.9	306	18	AAW38179	Arabidopsis SCARC
34	507	15.9	306	21	AAW28570	Arabidopsis SCla4.
35	506	15.8	306	21	AAW28598	Protein encoded by
36	498	15.6	480	21	AAW21049	Arabidopsis thalia
37	498	15.6	578	21	AAW21048	Arabidopsis thalia
38	488	15.6	593	21	AAW21047	Arabidopsis thalia
39	464	14.5	352	21	AAW28596	Protein encoded by
40	462.5	14.5	323	21	AAW28603	Maize CBPPT4 and
41	458	14.3	352	18	AAW38181	Arabidopsis SCARC
42	448	14.0	113	21	AAW32809	Eucalyptus grandis
43	443	13.9	809	21	AAW28586	Arabidopsis SCla2
44	441	13.8	808	21	AAW28602	Protein encoded by
45	439	13.7	541	21	AAW50885	Arabidopsis thalia

#### ALIGNMENTS

RESULT 1	
ID	AAV02540 standard; Protein; 623 AA.
XX	
AC	AAV02540;
XX	
DT	16-JUL-1999 (first entry)
XX	
DE	Protein encoded by wheat Rht clone 5a1 genomic sequence.
XX	
KW	Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;
KW	antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;
KW	paclobutrazol.
XX	
OS	Triticum aestivum.
XX	
PN	WO9909174-A1.
XX	
PD	25-FEB-1999.
XX	
PF	07-AUG-1998; 98WO-GB02383.
XX	
PR	13-AUG-1997; 97GB-0017192.
XX	
PA	(PLAN-) PLANT BIOSCIENCE LTD.
XX	
PI	Harberd NP, Peng J, Richards DE;
XX	
DR	WPI; 1999-181040/15.
XX	
DR	N-PSDB; AAW36279.
XX	
PT	New Triticum aestivum polynucleotides - encode a polypeptide which
PT	provides inhibition of the growth of plants, which inhibition is
PT	antagonised by gibberellin, used to confer a dwarf phenotype



```

|||||
Db 235 dtgeagirlvhalacaeavqgenfsaaalvkqjpmlasssggmrvkaayfgealarr 294
QY 286 VFRFRPODSSLLDAARFADLHAHFYESCPTLKFAHFTANQAILAEAFACRRVHVVDGCI 345
Db 295 YVFRPPDSSLLDAAFADLHAHFYESCPTLKFAHFTANQAILAEAFACRRVHVVDGCI 354
QY 346 KOGMOWPALLQALARPGGPSFRLTGVGPPDPDFTDALQOYGMKLAOPAHRTIRVDFQYR 405
Db 355 Kqgmwppallqalarrpgppsfritlyvpppdeidaqgvwklagfahrlirvdfqyr 414
QY 406 GLVAAATLADLEPFMLQPGSEDDPNEPEVIAVNSVFEMHRLAQGALAEKVLGVRAVAP 465
Db 415 glvaatladiapfmqlpgag-ddtddepevianvsfelhrllagpalekvlgtvrvavp 473
QY 466 RIVVVEOPAHNNSCTFLDRFTESLHYSTWPMDSLEGSSGGSPSEVSSGAAPAAACT 525
Db 474 RIVVVEOPAHNNSCTFLDRFTESLHYSTWPMDSLEGSSGGSPSEVSSGAAPAAACT 529
QY 526 DOYMEVYLGRQICNVVACEGAEERHERHETLGOMNRRLGNAGFEVYHLSNMYKQASTLL 585
Db 530 dqymsevylygrqicnvacegaeerhettlgqwrsllygsqfapvhlgsnaykqastll 589
QY 586 ALFAGDGYKVEKEGCTLTGWHTRPLIATSAMRLA 621
Db 590 alfagdgryveekdgctltgwhtrpliatssarva 625

RESULT 3
AA02539
ID AAY02539 standard; Protein: 425 AA.
XX
AC AAY02539;
XX
DT 16-JUL-1999 (first entry)
XX
DE Protein encoded by the wheat Rht clone C15-1 cDNA sequence.
XX
KW Rht gene; homologue: Triticum aestivum; wheat; growth inhibition;
KW antagonist: gibberellin; dwarf phenotype; gibberellin biosynthesis;
KW paclobutrazol.
XX
OS Triticum aestivum.
XX
PN W09909174-A1.
XX
PD 25-FEB-1999.
XX
PE 07-AUG-1998; 98WO-GB02383.
XX
PR 13-AUG-1997; 97GB-0017192.
XX
PA (PLAN-) PLANT BIOSCIENCE LTD.
XX
PI Harberd NP, Peng J, Richards DE;
XX
DR WPI: 1999-181040/15.
DR N-PSDB: AAX36278.
XX
PT New Triticum Aestivum polynucleotides - encode a polypeptide which
PT provides inhibition of the growth of plants, which inhibition is
PT antagonised by gibberellin, used to confer a dwarf phenotype
XX
PS Disclosure: Fig 7b; 88pp; English.
XX
CC The specification describes polypeptides encoded by the Rht gene (and
CC its homologues) that, when expressed in Triticum Aestivum, inhibit
CC growth of the plant. This growth inhibition is antagonised by
CC gibberellin. The products can be used to provide Rht expression in
CC plants, conferring a dwarf phenotype on a plant which is correctable
CC by treatment with gibberellin. In addition, the products can be
CC used to produce Rht mutant plants which are dwarfed compared with
CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants

```

CC may be made by knocking out Rht or the relevant homologous gene in  
 CC the plant of interest. Plants may be made which are resistant to  
 CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,  
 CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds  
 CC dwarf but let crop plants grow tall. The present sequence is encoded by  
 CC the wheat Rht clone C15-1 cDNA sequence.

XX Sequence 425 AA:

Query Match 68.5%; Score 2189; DB 20; Length 425;  
 Best Local Similarity 100.0%; Pred. No. 7.3e-165;  
 Matches 425; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 199 ARSSVEAAPVPAANAATPALPVPVYVDTQEGIRLVHALLKCAEVQDENSAAALVK 258
Db 1 arssvveaapvpaaanaatpalpvpvvdqceagirlvhalacaeavqgenfsaaalvk 60
QY 259 QIPLLAASQGAMRKVAAVFGELARVFRFRPODSSLLDAAFADLHAHFYESCPTLK 318
Db 61 qipllaasqsgamrvkaayfgealarrvfrfrpposslldaafadlhaHFYescpylk 120
QY 319 FAHFTANQAILAEAFACRRVHVVDGCIKOGMOWPALLQALARPGGPSFRLTGVGPPQ 378
Db 121 fahftangailaeafagcrrvhvvdfigkgmqwppallqalarrpgppsfritlyvppqp 180
QY 379 DETDALQOYGMKLAOPAHRTIRVDFQYRGLVAATLADLEPFMLQPGSEDDPNEPEVIAVN 438
Db 181 detdalqygvwklagfahrlirvdfqyrglvaatladielpfmjpegeedpnepevianv 240
QY 439 SVEEMHRLAQGALAEKVLGTAVRVRPIVTVVEOPAHNNSCTFLDRFTESLHYSTWMD 498
Db 241 svfemhrliaqgalekvlgtvrvavprivtvegeaahnsctfldrfeslhytctmd 300
QY 499 SLEGSSGGSPSEVSSGAAPAAAGTDQVMEVYLGRQICNVVACEGAEERHERHETLQ 558
Db 301 slegssggspsevssgaapaaagtdqymsevylygrqicnvacegaeerhettlq 360
QY 559 WNRRLGNAGFEVYHLSNMYKQASTLLAPAGDGYKVEKEGCTLTGWHTRPLIATSAM 618
Db 361 wnrllgnagfevhlgsnaykqastlllafagdykveekegctltgwhtrpliatssaw 420
QY 619 RLGP 623
Db 421 rlgap 425

RESULT 4
AAE02560
ID AAE02560 standard; Protein: 587 AA.
XX
AC AAE02560;
XX
DT 10-AUG-2001 (first entry)
XX
DE A. thaliana transcription factor G308 homolog, G307.
XX
KW plant transcription factor; phenotype; sugar sensing characteristic;
KW transgenic plant; plant yield; growth; germination; photosynthesis;
KW glyoxylate metabolism; respiration; pathogen response; wounding response;
KW cell cycle regulation; pigmentation; flowering; senescence; physiology;
KW storage organ; metabolism.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT Domain 323..339
FT /note="Conserved domain"
XX
PN W0200135725-A1.
XX
PD 25-MAY-2001.

```





DR WPI: 2001-335999/35.  
 DR N-PSDB: AAD05776.  
 PT Nucleic acids encoding plant transcription factor polypeptides, useful  
 PT for altering the biochemical characteristics of plants e.g. corn,  
 PT potato and cotton plants -  
 XX  
 PS Claim 4; Page 68-70; 127pp; English.  
 CC The present sequence is Arabidopsis thaliana transcription factor,  
 CC G307. The transcription factor is used for altering a plant's  
 CC biochemical characteristics. The transcription factor may be used to  
 CC alter the structure and developmental characteristics of plants such as  
 CC soybean, wheat, corn, potato, cotton, rice, oilseed rape, sunflower,  
 CC alfalfa, sugar cane, turf, banana, blackberry, blueberry, strawberry,  
 CC raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant,  
 CC grapes, honey dew, lettuce, mango, melon, onion, papaya, peas, peppers,  
 CC pineapple, spinach, squash, sweet corn, tobacco, tomato, watermelon,  
 CC rosaceous fruits and/or vegetable brassicas. Transcription factors are  
 CC key controlling elements of biological pathways and altering expression  
 CC levels of 1 or more transcription factors can change entire biological  
 CC pathways in an organism. Therefore manipulating transcription factor  
 CC levels in plants offers great potential in agricultural biotechnology  
 CC for modifying a plant's traits. Transcription factor cDNA is useful in  
 CC gene therapy.  
 CC  
 XX  
 XX Sequence 587 AA:

Query Match 50.5%; Score 1612; DB 22; Length 587;  
 Best Local Similarity 53.7%; Pred. No. 4,7e-119;  
 Matches 340; Conservative 82; Mismatches 149; Indels 62; Gaps 13;

QY 1 MKREYQDAGG---SGGGGGGSGSSEDKMMV--SAAAGEEVEDLLAAGYKVRASDNA 54  
 DB 1 mkrdhgfgrlsmhgtsessssiskmmvmvkkeedggmndellavlygkvrssema 60  
 QY 55 DVAOKLEOLEMAMGAGVAGAAPDDSFATHATDTVHNPNPDLSSWVSMSLNAAPP 114  
 DB 61 evalleqletmm-----snvedglshlatlvhyhpelyswldnmseIn--pp 110  
 QY 115 PTPAPQLMASTSTVYTGSGYFDLPSPVD--SSSIALPRIPSPAGATPA--DLSD 170  
 DB 111 ptp-----aasnqldpvlpspeicgfpesdydkvlpnaalyqfpadssss 157  
 QY 171 SVRDKRMRTGSSST--SSSSSSSLGSGARSVVEAAPVAANAATPALPVVVVDRO 228  
 DB 158 smngqkrlkscspdsmyststgtqigvltvtltttttaaesfrs--vllvdsq 215  
 QY 229 ENGIRLVNALLACAAVQOENLSAABALVKQIPLAASGCGAMRKVAAYFGEALARRVR 288  
 DB 216 engvrlvnlhmeaaalqgnltlaeaalvkqigclavsqagamrkvaayfaealarrry 275  
 QY 289 FRPOPDSLLDAAFDLTAHFESCPYTKFAHTANQAILLEAFACRVRHVVDFIGIKG 348  
 DB 276 l--spqqngidhclsdltqbmhyecopylkfahfctangallaeegkkrvvhidfdmng 333  
 QY 349 MQWPALLQALALPGGSPSEFRITGVGPDPDETDALQGVGMKLAQFAHTIRVDFOYRGIV 408  
 DB 334 lqwpalmqalalregppfrfritgtpgpadnsdhlyevcklaqaaehvefeyrgiv 393  
 QY 409 AATLADLEPFMLQPGCEENPEPEYIANVNSYFEMHRLAOGALEKVLGYTRAVNPRIV 468  
 DB 394 anlsladldasml-----elrps--dteavavsvfelhklilgpgglekvlgvqqkllpvl 448  
 QY 469 TVVEDEANNSGTFIDRFESLHYSTFMDPSLEGGSSGGSPSEVSSGAAPAAACTDOV 528  
 DB 449 tvvegesnmgndfdrfeshlystfmdpsle-----vpsqdkv 490  
 QY 529 MSEVYLGROIQNVNACEGARTERHETTLQOMNRNLGNAGFETVHLGSANVYKQASTLLALF 588  
 DB 491 msevylgkroicnvacegdrverhetlsgwngtrfgsgslapahgnsafkqasmllsvf 550

QY 589 AGDDGYKEEGCLTGLMHTPLIATSAMRLA 621  
 DB 551 msqgyrveesngclmghwtrpllttsawkls 583

RESULT 6  
 AAM30792  
 ID AAM30792 standard; Protein; 532 AA.

AC AAM30792;  
 DT 19-MAR-1998 (first entry)

DE Arabidopsis thaliana gibberellin insensitivity gai gene product.  
 XX  
 KW gibberellin insensitivity; gai; plant growth inhibition;  
 KW dwarf phenotype; lodging resistance; increased yield;  
 KW flowering regulation; bolting inhibition; spinach; lettuce;  
 KW antibody; identification; probe; primer; antisense; sense;  
 KW expression regulation; co-suppression; rice;  
 KW Bakane disease resistance.

OS Arabidopsis thaliana.

PN WO9729123-A2.

PD 14-AUG-1997.

PF 12-FEB-1997; 97WO-GB00390.

PR 12-FEB-1996; 96GB-0002796.

PA (INNE-) INNES CENT INNOVATIONS LTD JOHN.

DR Carol P, Harberd NP, Peng J, Richards DE;

XX WPI: 1997-415295/38.

DR N-PSDB: AAT91937.

PT Nucleic acid encoding gibberellin inhibitor GAI and related  
 PT antisense sequences - used to create tall, or particularly, dwarf  
 PT plants, especially crops such as maize, rice and wheat

PS Claim 1; Fig 4: 76pp; English.

XX The present sequence is the Arabidopsis thaliana gibberellin  
 CC insensitivity (gai) gene product (GAI), the expression of which  
 CC inhibits plant growth. However the inhibition is antagonised by  
 CC gibberellin (GA), while gai expression confers a dwarf phenotype  
 CC that is insensitive to GA. Manipulating gai and GAI expression can  
 CC produce tall or dwarf plants, particularly the latter for increased  
 CC resistance to lodging and increased yield. It may also allow  
 CC regulation of flowering, i.e. plants remain in the vegetative state  
 CC until treated with GA, useful to inhibit bolting in spinach and  
 CC lettuce. GAI can be used to raise specific antibodies for  
 CC identifying homologous proteins or genes in other species. gai  
 CC fragments can also be used as probes or primers to identify and  
 CC clone related sequences, or in the preparation of antisense or  
 CC sense expression regulating (co-suppressing) sequences. Rice plants  
 CC that express GAI may be resistant to Bakane disease. Manipulation  
 CC of gai and GAI makes it possible to tailor the degree of dwarfism  
 CC and GA sensitivity to particular crops or situations.

XX Sequence 532 AA:

Query Match 49.9%; Score 1594; DB 18; Length 532;  
 Best Local Similarity 54.3%; Pred. No. 1.1e-117;  
 Matches 332; Conservative 71; Mismatches 106; Indels 102; Gaps 11;

QY 22 EDMMVSAAGEEVEDLLAAGYKVRASDMADVAKLEOLEMAMGAGVAGAAPDD 81  
 DB 11 dktmmneedgngmdellavlygkvrssmadvqkleglevmm-----snvedd- 63



[illegible]

RESULT 8  
AAE02545  
ID AAE02545 standard; Protein: 532 AA.  
XX  
AC AAE02545;  
XX  
D7 10-AUG-2001 (first entry)  
XX  
DE A. thaliana transcription factor G308.  
XX  
KW Plant transcription factor; phenotype; sugar sensing characteristic;  
KW transgenic plant; plant yield; growth; germination; photosynthesis;  
KW glyoxylate metabolism; respiration; pathogen response; wounding response;  
KW cell cycle regulation; pigmentation; flowering; senescence; physiology;  
KW storage organ; metabolism.  
XX  
OS Arabidopsis thaliana.  
XX  
FH Key Location/Qualifiers  
FT Domain 270..274  
FT /note="Conserved domain"  
XX  
PD MO200135725-A1.  
XX  
PD 25-MAY-2001.  
XX  
PF 14-NOV-2000; 2000MO-US31414.  
XX  
PR 17-NOV-1999; 99US-0166228.  
PR 17-APR-2000; 2000US-0197899.  
PR 22-AUG-2000; 2000US-0227439.  
XX  
PA (MEND-) MENDEL BIOTECHNOLOGY INC.  
PA (JIAN-) JIANG C.  
PA (HEAR-) HEARD J.  
PA (PINE-) PINEDA O.  
PA (PILG-) PILGRIM M.  
PA (ADAM-) ADAM L.  
PA (RIEC-) RIECHMANN J L.  
PA (YUGG-) YU G.  
PA (SAMA-) SAMAH R.  
PI Jiang C, Heard J, Plineda O, Pilgrim M, Adam L, Riechmann JL;  
PI Yu G, Samaha R;  
PI WPI: 2001-335977/35.  
DR N-PSDB: AAD06646.  
XX  
XX Nucleic acids encoding plant transcription factor polypeptides, useful  
PT for altering the sugar sensing characteristics of plants and increasing  
PT yield, e.g. corn, potato and cotton plants -  
XX  
XX Claim 4; Page 74-76; 151pp; English.

CC onion, papaya, peas, peppers, pineapple, spinach, squash, sweet corn,  
CC tobacco, tomato, watermelon, rosaceous fruits and/or vegetable brassicas.  
CC The present sequence is an Arabidopsis thaliana transcription factor.  
XX  
SQ Sequence 532 AA:  
  
Query Match 49.8%; Score 1590; DB 22; Length 532;  
Best Local Similarity 54.2%; Pred. No. 2,2e-117;  
Matches 331; Conservative 72; Mismatches 106; Indels 102; Gaps 11:  
  
OY 22 EDKMWVSAAGBGEVEDELLAALGKVRASDMADVAKLEOLEMAMGCGAGAAPDS 81  
DB 11 dkktmmneedgngmdellavlygkvrsemadvakleqlevnm-----snvqedd- 63  
OY 82 FATHATDTVHNPRDLSWVESMLSELNAPPRLPPAPQALNASSVTGSGGFDDLP 141  
DB 64 -lsqlatetvhympaeltywldsmldinp----- 93  
OY 142 SVDSSSIYALRPIPSAGATPAPADLSADSVDPKMRGTSGSSSSSSSLGCG--- 198  
DB 94 ---snaeydikaip-----gdall-----ngfaidaassngsggddt 129  
OY 199 -----ARSSVEADPPVAAAANATPALPVVVDTQEGIRLVHALLACAEAVQENL 250  
DB 130 ytnkrlkcsngvve-----ttlaesttrhvlyvdsqengyrlvhallicaeavqenl 184  
OY 251 SAAELVQIPIPLAASOGGAMKVAAYGCEALARVFRFRPQDSLLUDAAFDLHMF 310  
DB 185 lvaalvqxqigflavsqigamqvallylaealarlyrl--spsgspdhslsdlqmf 242  
OY 311 YESCPYLFKFAHTANQALILEAFAGCRVHVVDGIGKGMQMPALQALALRGPSPSRL 370  
DB 243 yelcplylkfahltanqalaleagkkrvhvldfsmgslqwpalngalalrpggprvrl 302  
OY 371 TGVGPPQPEDTDALQOVGKMLAQFAHTIRVDFOYRGVLAATLADLEPMLDPEGSEDRNE 430  
DB 303 tgisppapdnfdylyevgcklahlaeahvefeyrgfvantladdaaml---elrps- 357  
OY 431 EPEVFAVNSVFEHRLILQPGALEKLVTRAVRPRITVYVQEANHNSGTFLLDFTSL 490  
DB 358 elsvavnsvfeihkllgrpaldkvlgvnglkpeellvveqeshnspflidrfteei 417  
OY 491 HYSTMPDLSLEGSSGGSPSEVSSGAAAAPAAAGTDQVMSVYTLGROIQNVVACGAERT 550  
DB 418 hystlfdslsg-----ypsgqdkvmseylygkqlcnvvaacdprtv 459  
OY 551 ERHETLGGMRNRLGNAGFETVHLGSNATKQASTLLALFAGDNGYVEKEGCLTGHTR 610  
DB 460 ethetlsgwtrnrfgsagfaaahlgnaikqasmllalngesgyveesdgclmlywhlr 519  
OY 611 PLIANTSAMRLA 621  
DB 520 pliatcawklis 530  
  
RESULT 9  
AAE01907  
ID AAE01907 standard; Protein: 532 AA.  
XX  
AC AAE01907;  
XX  
D7 31-JUL-2001 (first entry)  
XX  
DE Arabidopsis thaliana transcription factor, G308.  
XX  
KW Transcription factor; biochemical characteristic; controlling element;  
KW structural characteristic; developmental characteristic; gene therapy;  
KW agricultural biotechnology; plant trait modification.  
XX  
OS Arabidopsis thaliana.  
XX  
FH Key Location/Qualifiers

Domain 270..274  
/Label= Conserved\_domain

WT Domain 270..274  
XX /Label= Conserved\_domain

XX MO200136597-A1.

XX 25-MAY-2001.

XX 14-NOV-2000; 2000WO-US31344.

XX 17-NOV-1999; 99US-0166228.

XX 17-APR-2000; 2000US-0197899.

XX 22-AUG-2000; 2000US-0227439.

XX (MEMD-) MENDEL BIOTECHNOLOGY INC.

XX (GREG/) CREELMAN R.

XX (YUGG/) YU G.

XX (ADAM/) ADAM L.

XX (RIEC/) RIECHMANN J L.

XX (HEAR/) HEARD J.

XX (SAMA/) SAMARA R.

XX (PIUG/) PILGRIM M.

XX (PINE/) PINEDA O.

XX (JIANG/) JIANG C.

XX Creelman R, Yu G, Adam L, Riechmann JL, Heard J, Samara R;  
PI Pilgrim M, Pineda O, Jiang C;  
XX WPI: 2001-335999/35.  
XX N-PSDB: AAD05791.

XX Nucleic acids encoding plant transcription factor polypeptides, useful  
PT for altering the biochemical characteristics of plants e.g. corn,  
PT potato and cotton plants -

XX Claim 4; Page 114-115; 127pp; English.

XX The present sequence is Arabidopsis thaliana transcription factor,  
CC G308, a homologue of G307. The transcription factor is used for altering  
CC a plant's biochemical characteristics. The transcription factor may be  
CC used to alter the structure and developmental characteristics of plants  
CC such as soybean, wheat, corn, potato, cotton, rice, oilseed rape,  
CC sunflower, alfalfa, sugar cane, turf, banana, blackberry, blueberry,  
CC strawberry, raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber,  
CC eggplant, grapes, honey dew, lettuce, mango, melon, onion, papaya, peas,  
CC peppers, pineapple, spinach, squash, sweet corn, tobacco, tomato,  
CC watermelon, rosaceous fruits and/or vegetable brassicas. Transcription  
CC factors are key controlling elements of biological pathways and altering  
CC expression levels of 1 or more transcription factors can change entire  
CC biological pathways in an organism. Therefore manipulating transcription  
CC factor levels in plants offers great potential in agricultural  
CC biotechnology for modifying a plant's traits. Transcription factor cDNA  
CC is useful in gene therapy.

XX Sequence 532 AA:

Query Match 49.8%; Score 1590; DB 22; Length 532;  
Best Local Similarity 54.2%; Pred. NO. 2.2e-117;  
Matches 331; Conservative 72; Mismatches 106; Indels 102; Gaps 11;

QY 22 EKKMVAASAAAGGEEVDELALALGYVRASDMADVAQKIEQLEMGMGCVGAGAAPDD 81  
DB 11 dkkmmmeedgngmdellavlygvrrsemadvqkleglevmm-----snvgedd- 63  
QY 82 FATHTLATDVHNPPTDLSSVSMSELNAPPPLPPAPQALNASTSVTSGGYFDLP 141  
DB 64 -lsqlatevhnpeelywmidsmltdlmp----- 93  
QY 142 SYDSSSYIALRPISPPAGATAPADISADVNDPKRMRTGSSSTSSSSSSSLGSG--- 198  
DB 94 ---ssnaeydlkaip-----gdaij-----ngfaidsasssnqgsgygt 129  
QY 199 -----ARSSVVEAPPAVAAANATPALPVVVDTQENGRITLVALLACAEAVQENL 250

DB 130 yctnrklcsngvve-----ttataestrhvlvdsgngvrylvhallaceavqkenl 184  
QY 251 SAAEALVKQIPLLAASOGGAMRKVAAYFGEALARRFRFPQDSSLDAAPDLHAHF 310  
DB 185 tvaeealvkqifglavsqigamqvallyfaealarlyrl--spsgpishsldtlgmhf 242  
QY 311 YESCPYKFAHFANQALIEAFAGCRHVVDGFIKQGMQWALLQALALRQGPSPFRL 370  
DB 243 yetcpylkfahtanaqalleatgqkrvnhldfsmgsqilqwpalmqalalrpgppvftl 302  
QY 371 TGVGPQPDETDALQVGMKLAOFATHTINVDFOYRGLVATLADLEPFMIQEGEEDPNE 430  
DB 303 tglpppagnfdytlevgaklahlaeeahvefevrgfvantladiasl----elrps- 357  
QY 431 EPEVIAVNSVFEMKRLAOPGALKVLTGVRAPRIYTVROEANHNGSTFLDRTESL 490  
DB 358 elesvavnsvfelkhllygpaidkylgvnqkpelftvveqeshnspifldrtfesi 417  
QY 491 HYSTMFDSLSESSSGGSPSEVSSGAAPAAAGTDQVMSSEVYLGROICNVVACGEAERT 550  
DB 418 hystlfdslleg-----vpsgqdkvmsvlylgqlcnvvaedgpdtr 459  
QY 551 EHEETLGMWRNLGNAGFEYVHLGSNAYRQASTLLALFAGDGYKYEKEGCTLLGMHTR 610  
DB 460 ehetslsgwrntfsgsfafaahgnaafkqaemllalfnggegyrveesdcmlgmwhtr 519  
QY 611 PLIANSAMRLA 621  
DB 520 platsawklis 530

RESULT 10  
AAG38577  
ID AAG38577 standard; Protein; 517 AA.  
XX  
AC AAG38577;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 47612.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PE 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0128745.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132407.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140365.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 16-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 19-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147306.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 27-AUG-1999; 99US-0151303.  
PR 30-AUG-1999; 99US-0151330.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157755.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159299.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.













---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2002, 16:24:19 ; Search time 22.72 Seconds  
(without alignments)  
12.876 Million cell updates/sec

Title: US-09-485-529-103  
Perfect score: 76  
Sequence: 1 LNAPPPPLPPAPQ 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents\_AA: \*  
1: /cgn2\_6/prodata/2/iaa/5A\_COMB.pep: \*  
2: /cgn2\_6/prodata/2/iaa/5B\_COMB.pep: \*  
3: /cgn2\_6/prodata/2/iaa/6A\_COMB.pep: \*  
4: /cgn2\_6/prodata/2/iaa/6B\_COMB.pep: \*  
5: /cgn2\_6/prodata/2/iaa/PCTUS\_COMB.pep: \*  
6: /cgn2\_6/prodata/2/iaa/backfilest1.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	55	72.4	446	1	US-07-952-800-4
2	54	71.1	3118	2	US-08-457-273B-8
3	53	69.7	726	4	US-09-126-980-2
4	53	69.7	726	4	US-09-126-980-2
5	52	68.4	15	2	US-08-769-745-31
6	52	68.4	281	2	US-08-810-453-2
7	52	68.4	281	3	US-08-815-190A-2
8	52	68.4	281	4	US-09-290-640-25
9	52	68.4	281	4	US-09-479-524-3
10	52	68.4	281	5	PCT-US95-00362-2
11	52	68.4	288	3	US-08-545-196B-21
12	52	68.4	288	4	US-09-028-327-3
13	52	68.4	311	4	US-09-179-558-66
14	52	68.4	311	4	US-09-179-558-66
15	51	67.1	32	3	US-08-545-196B-63
16	51	67.1	170	5	PCT-US95-16806A-6
17	51	67.1	278	3	US-08-545-196B-19
18	51	67.1	294	1	US-08-612-986-6
19	51	67.1	294	1	US-08-361-806A-6
20	51	67.1	294	3	US-08-545-196B-9
21	51	67.1	294	4	US-09-028-327-4
22	51	67.1	294	5	PCT-US95-16806A-21
23	51	67.1	445	2	US-08-900-148-2
24	50	65.8	15	4	US-08-602-999A-430
25	50	65.8	15	4	US-08-602-999A-431
26	50	65.8	18	4	US-08-602-999A-371
27	50	65.8	20	4	US-08-602-999A-368

28	50	65.8	22	4	US-08-602-999A-448	Sequence 448, App
29	50	65.8	408	2	US-08-926-258-2	Sequence 2, App1
30	50	65.8	408	2	US-09-120-053-2	Sequence 2, App1
31	50	65.8	408	4	US-09-198-212-2	Sequence 2, App1
32	50	65.8	525	4	US-09-041-886-28	Sequence 28, App1
33	50	65.8	525	4	US-08-764-870-7	Sequence 7, App1
34	50	65.8	525	4	US-08-980-115-7	Sequence 29, App1
35	50	65.8	530	4	US-09-041-886-29	Sequence 2, App1
36	50	65.8	533	1	US-07-952-800-2	Sequence 30, App1
37	50	65.8	552	4	US-09-041-886-30	Sequence 31, App1
38	50	65.8	589	4	US-09-041-886-31	Sequence 16, App1
39	50	65.8	3119	1	US-08-246-982A-16	Sequence 16, App1
40	50	65.8	3119	1	US-08-453-265-16	Sequence 6, App1
41	50	65.8	3144	1	US-08-246-982A-6	Sequence 6, App1
42	50	65.8	3144	1	US-08-453-265-6	Sequence 42, App1
43	50	65.8	3144	2	US-08-457-273B-42	Sequence 21, App1
44	50	65.8	3144	3	US-08-556-419-21	Sequence 15, App1
45	50	65.8	3144	4	US-09-041-886-15	

## ALIGNMENTS

RESULT 1  
US-07-952-800-4  
; Sequence 4, Application US/07952800  
; Patent No. 5403925  
; GENERAL INFORMATION:  
; APPLICANT: OZATO, KEIKO  
; TITLE OF INVENTION: A NEW MEMBER OF THE NUCLEAR HORMONE  
; TITLE OF INVENTION: RECEPTOR SUPERFAMILY AND A CDNA CLONE THEREOF  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TOWNSEND AND TOWNSEND  
; STREET: 1 MARKER PLAZA, STEWART TOWER, SUITE 2000  
; CITY: SAN FRANCISCO  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94105  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/952,800  
; FILING DATE: 19920928  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Weber, Kenneth A  
; REGISTRATION NUMBER: 31,677  
; REFERENCE/DOCKET NUMBER: 15280-21-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-543-9600  
; TELEFAX: 415-543-5043  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 446 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-07-952-800-4

Query Match 72.4%; Score 55; DB 1; Length 446;  
Best Local Similarity 72.7%; Pred. No. 10;

Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 2 NAPPPLPPAP 12  
DB 31 SAPPPLPPAP 41

```
RESULT 2
US-08-457-273B-8
; Sequence 8, Application US/08457273B
; Patent No. 5849995
; GENERAL INFORMATION:
; APPLICANT: Hayden, Michael
; APPLICANT: Lin, Biaoyang
; APPLICANT: Nasir, Jamal
; TITLE OF INVENTION: Mouse Model for Huntington's Disease and
; TITLE OF INVENTION: Related DNA Sequences
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESS: Virginia Bennett
; STREET: PO Box 37428
; CITY: Raleigh
; STATE: No. 5849995th Carolina
; COUNTRY: US
; ZIP: 27627
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,273B
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Bennett, Virginia C.
; REGISTRATION NUMBER: 37,092
; REFERENCE/DOCKET NUMBER: 3477-85A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-854-1400
; TELEFAX: 919-854-1401
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3118 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-457-273B-8

Query Match          71.1%; Score 54; DB 2; Length 3118;
Best Local Similarity 81.8%; Pred. No. 83;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 APPPLPAPQ 13
||||| 1111
Db 29 APPPPPPPPQ 39

RESULT 3
US-09-126-980-2
; Sequence 2, Application US/09126980
; Patent No. 6270956
; GENERAL INFORMATION:
; APPLICANT: Jones, Katherine
; APPLICANT: Wei, Ping
; APPLICANT: Gardner, Mitchell
; APPLICANT: Fang, Shi-Min
; TITLE OF INVENTION: A TRANSCRIPTIONAL COACTIVATOR THAT
; TITLE OF INVENTION: INTERACTS WITH TAT PROTEIN AND REGULATES ITS
; TITLE OF INVENTION: BINDING TO TAR RNA, METHODS FOR MODULATING TAT
; TITLE OF INVENTION: TRANSCRIPTION, AND USES THEREFOR
; FILE REFERENCE: SALK2231
; CURRENT APPLICATION NUMBER: US/09/126,980
; CURRENT FILING DATE: 1998-07-30
; EARLIER APPLICATION NUMBER: 60/069,341
; EARLIER FILING DATE: 1997-12-11
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
; SEQ ID NO 2
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-126-980-2

Query Match          69.7%; Score 53; DB 4; Length 726;
Best Local Similarity 80.0%; Pred. No. 27;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 PPPPLPAPQ 13
||||| 1111
Db 717 PPPPLPPLPK 726

RESULT 4
US-09-476-482-2
; Sequence 2, Application US/09476482
; Patent No. 6284456
; GENERAL INFORMATION:
; APPLICANT: Jones, Katherine A.
; APPLICANT: Wei, Ping
; APPLICANT: Gardner, Mitchell
; APPLICANT: Fang, Shi-Min
; TITLE OF INVENTION: A TRANSCRIPTIONAL COACTIVATOR THAT
; TITLE OF INVENTION: INTERACTS WITH TAT PROTEIN AND REGULATES ITS BINDING TO TAR
; TITLE OF INVENTION: RNA, METHODS FOR MODULATING TAT TRANSCRIPTION, AND USES
; FILE REFERENCE: SALK2230-2
; CURRENT APPLICATION NUMBER: US/09/476,482
; CURRENT FILING DATE: 1999-12-30
; EARLIER APPLICATION NUMBER: 09/126,980
; EARLIER FILING DATE: 1998-07-30
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-476-482-2

Query Match          69.7%; Score 53; DB 4; Length 726;
Best Local Similarity 80.0%; Pred. No. 27;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 PPPPLPAPQ 13
||||| 1111
Db 717 PPPPLPPLPK 726

RESULT 5
US-08-769-745-31
; Sequence 31, Application US/08769745
; Patent No. 5955259
; GENERAL INFORMATION:
; APPLICANT: Holmes, Todd C.
; APPLICANT: Levitan, Todd B.
; APPLICANT: Brandeis University
; TITLE OF INVENTION: Mechanism for the Regulation of Ion
; FILE REFERENCE: BRU96-02
; CURRENT APPLICATION NUMBER: US/08/769,745
; CURRENT FILING DATE: 1996-12-19
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 31
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Homo sapien
US-08-769-745-31
```

Query Match 68.4%; Score 52; DB 2; Length 15;  
Best Local Similarity 88.9%; Pred. No. 0.94;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 PPPPLPPAP 12  
| | | | | | | |  
DB 3 PPPPLPPPP 11

RESULT 6  
US-08-810-453-2

; Sequence 2, Application US/08810453  
; Patent No. 5858990  
; GENERAL INFORMATION:  
; APPLICANT: Walsh, Kenneth  
; TITLE OF INVENTION: FAS LIGAND COMPOSITIONS FOR TREATMENT OF  
; TITLE OF INVENTION: PROLIFERATIVE DISORDERS  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
; STREET: Federal Reserve Plaza, 600 Atlantic Avenue  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02210-2211  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/810,453  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plumber, Elizabeth R.  
; REGISTRATION NUMBER: 36,637  
; REFERENCE/DOCKET NUMBER: S1237/7004  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)720-3500  
; TELEFAX: (617)720-2441  
; TELEX: 343  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 281 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-810-453-2

Query Match 68.4%; Score 52; DB 2; Length 281;  
Best Local Similarity 88.9%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 PPPPLPPAP 12  
| | | | | | | |  
DB 50 PPPPLPPPP 58

RESULT 7  
US-08-815-190A-2

; Sequence 2, Application US/08815190A  
; Patent No. 6046310  
; GENERAL INFORMATION:  
; APPLICANT: Queen, Cary L.  
; APPLICANT: Schneider, William P.  
; APPLICANT: Vasquez, Maximiliano  
; TITLE OF INVENTION: Fas Ligand Fusion Proteins and their  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/815,190A  
; FILING DATE: 11-MAR-1997  
; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/614,584  
; FILING DATE: 13-MAR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Apple, Randolph T.  
; REGISTRATION NUMBER: 36,429  
; REFERENCE/DOCKET NUMBER: 011823-006710US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 281 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-815-190A-2

Query Match 68.4%; Score 52; DB 3; Length 281;  
Best Local Similarity 88.9%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 PPPPLPPAP 12  
| | | | | | | |  
DB 50 PPPPLPPPP 58

RESULT 8

US-09-290-640-25  
; Sequence 25, Application US/09290640  
; Patent No. 6204055  
; GENERAL INFORMATION:  
; APPLICANT: Marcusson, Eric G.  
; APPLICANT: Dean, Nicholas M.  
; TITLE OF INVENTION: Antisense Compound Modulation of Fas Mediated Signaling  
; FILE REFERENCE: ISPH-0351  
; CURRENT APPLICATION NUMBER: US/09/290,640  
; CURRENT FILING DATE: 1999-04-12  
; NUMBER OF SEQ ID NOS: 85  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 25  
; LENGTH: 281  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-290-640-25

Query Match 68.4%; Score 52; DB 4; Length 281;  
Best Local Similarity 88.9%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 PPPPLPPAP 12  
| | | | | | | |  
DB 50 PPPPLPPPP 58

RESULT 9

US-09-479-524-3  
; Sequence 3, Application US/09479524  
; Patent No. 6268350  
; GENERAL INFORMATION:  
; APPLICANT: Barbara Gulliem, Emilio  
; TITLE OF INVENTION: Polynucleotides For Inhibiting Metastasis And Tumor Cell Growth  
; FILE REFERENCE: B-37  
; CURRENT APPLICATION NUMBER: US/09/479,524  
; CURRENT FILING DATE: 2000-01-07  
; EARLIER APPLICATION NUMBER: US 09/170,948  
; EARLIER FILING DATE: 1998-10-13  
; EARLIER APPLICATION NUMBER: US 60/062,733  
; EARLIER FILING DATE: 1997-10-23  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: Wordperfect  
; SEQ ID NO: 3  
; LENGTH: 281  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-479-524-3

Query Match 68.4%; Score 52; DB 4; Length 281;  
Best Local Similarity 88.9%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 PPPPLPPAP 12  
| | | | | | |  
Db 50 PPPPLPPPP 58

RESULT 10  
PCT-US95-00362-2  
; Sequence 2, Application PC/TUS9500362  
; GENERAL INFORMATION:  
; APPLICANT: IMMUNEX CORPORATION  
; TITLE OF INVENTION: Ligand That Binds Fas Antigen  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Immunex Corporation  
; STREET: 51 University Street  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: US  
; ZIP: 98101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Apple Macintosh  
; OPERATING SYSTEM: Apple 7.1  
; SOFTWARE: Microsoft Word, Version 5.1a  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/00362  
; FILING DATE: 06-JAN-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/179,138  
; FILING DATE: 07-JAN-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/190,559  
; FILING DATE: 01-FEB-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Anderson, Kathryn A.  
; REGISTRATION NUMBER: 32,172  
; REFERENCE/DOCKET NUMBER: 2805-WO  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 281 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
PCT-US95-00362-2

Query Match 68.4%; Score 52; DB 5; Length 281;  
Best Local Similarity 88.9%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 PPPPLPPAP 12  
| | | | | | |  
Db 50 PPPPLPPPP 58

RESULT 11  
US-08-545-196B-21  
; Sequence 21, Application US/08545196B  
; Patent No. 6080577  
; GENERAL INFORMATION:  
; APPLICANT: MELKI, JUDITH  
; APPLICANT: MUNNICH, ARNOLD  
; TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE: A GENE  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP  
; STREET: PO BOX 747  
; CITY: FALLS CHURCH  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentln Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/545,196B  
; FILING DATE: 19-OCT-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FARACI, C. J.  
; REGISTRATION NUMBER: 32,350  
; REFERENCE/DOCKET NUMBER: 2121-110P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 205-8000  
; TELEFAX: (703) 205-8050  
; INFORMATION FOR SEQ ID NO: 21:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 288 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-545-196B-21

Query Match 68.4%; Score 52; DB 3; Length 288;  
Best Local Similarity 88.9%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 PPPPLPPAP 12  
| | | | | | |  
Db 214 PPPPLPPPP 222

RESULT 12  
US-09-028-327-3  
; Sequence 3, Application US/09028327  
; Patent No. 6130064  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Guegler, Karl J.  
; TITLE OF INVENTION: HUMAN SMN-LIKE PROTEIN  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Dr.



CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/028,327  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0482 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 288 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 1857114  
US-09-028-327-3

Query Match 68.4%; Score 52; DB 4; Length 288;  
Best Local Similarity 88.9%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 4 PPPPLPPAP 12  
|||||||  
Db 214 PPPPLPPPP 222

RESULT 13  
US-09-179-558-66  
; Sequence 66, Application US/09179558  
; Patent No. 6180612  
; GENERAL INFORMATION:  
; APPLICANT: Hockensmith, Joel W.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR  
; TARGETING DNA METABOLIC PROCESSES USING  
; TITLE OF INVENTION: AMINOGLYCOSIDE DERIVATIVES  
; NUMBER OF SEQUENCES: 66  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/179,558  
; FILING DATE: 27-OCT-1998  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: U.S. 09/060,470  
FILING DATE: 15-APR-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.S. 60/063,898  
FILING DATE: 31-OCT-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 9426-005-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)7909090  
TELEFAX: (212)8699741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 66:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 311 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-179-558-66

Query Match 68.4%; Score 52; DB 4; Length 311;  
Best Local Similarity 88.9%; Pred. No. 16;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 4 PPPPLPPAP 12  
|||||||  
Db 12 PPPPLPPPP 20

RESULT 14  
US-07-945-283-2  
; Sequence 2, Application US/07945283  
; Patent No. 5352596  
; GENERAL INFORMATION:  
; APPLICANT: Cheung, Andrew K.  
; APPLICANT: Wesley, Ronald D.  
; TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants  
; INVOLVING THE EP0 AND LIT GENES  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Curtis P. Ribando  
; STREET: 1815 No. 5352596th University Street  
; CITY: Peoria  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 61604  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/945,283  
; FILING DATE: 19920911  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ribando, Curtis P  
; REGISTRATION NUMBER: 27976  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 309-685-4011 ext. 513  
; TELEFAX: 309-685-4128  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1958 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-07-945-283-2

Query Match 68.4%; Score 52; DB 1; Length 1958;  
Best Local Similarity 88.9%; Pred. No. 91;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 PPPPLPPAP 12  
Db 489 PPPPLPPP 497

## RESULT 15

US-08-545-196B-63  
; Sequence 63, Application US/08545196B  
; Patent No. 6080577  
; GENERAL INFORMATION:  
; APPLICANT: MELKI, JUDITH  
; APPLICANT: MUNNICH, ARNOLD  
; TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE: A GENE  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP  
; STREET: PO BOX 747  
; CITY: FALLS CHURCH  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentln Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/545,196B  
; FILING DATE: 19-OCT-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FARACI, C. J.  
; REGISTRATION NUMBER: 32,350  
; REFERENCE/DOCKET NUMBER: 2121-110P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 205-8000  
; TELEFAX: (703) 205-8050  
; INFORMATION FOR SEQ ID NO: 63:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 32 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-545-196B-63

Query Match 67.1%; Score 51; DB 3; Length 32;  
Best Local Similarity 72.7%; Pred. No. 2.5;  
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 NAPPPLPPAP 12  
Db 6 NGPPPPPPPPP 16

Search completed: January 22, 2002, 16:24:20  
Job time: 40 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2002, 16:23:40 ; Search time 26.73 Seconds  
(without alignments)  
1775.411 Million cell updates/sec

Title: US-09-485-529-7

Perfect score: 3194

Sequence: 1 MKREYQDAGSGGGGGGSGMS.....TLGHTPLATSAWRLGP 623

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1608	50.3	587	2	D84426 hypothetical prote
2	1594	49.9	533	2	H86282 protein F10B6.34 [
3	1511.5	47.3	511	2	G96688 hypothetical prote
4	1450.5	45.4	523	2	T51475 RGA-like protein -
5	604.5	18.9	653	2	T47581 SCARECROW1 - Arabi
6	600.5	18.8	653	2	T51244 SCARECROW protein
7	572.5	17.9	482	2	E96540 hypothetical prote
8	558	17.5	526	2	E96542 SCARECROW-like pro
9	554	17.3	413	2	G84462 SCARECROW-like pro
10	551	17.3	658	2	D96656 hypothetical prote
11	512	16.0	313	2	T51242 Scil protein (limpo
12	507	15.9	306	2	T51236 SCARECROW-like pro
13	498	15.6	593	2	E86347 SCARECROW-like pro
14	464	14.5	352	2	T51234 SCARECROW-like pro
15	445.5	13.9	1336	2	T02736 probable SCARECROW
16	444.5	13.9	542	2	T46142 SCARECROW-like 7 (
17	443	13.9	808	2	T51232 SCARECROW-like pro
18	433.5	13.6	531	2	T04722 SCARECROW-like prote
19	429.5	13.4	284	2	T51241 SCARECROW-like pro
20	415	13.0	375	2	C71441 probable SCARECROW
21	402.5	12.6	325	2	T51235 SCARECROW-like pro
22	399.5	12.5	718	2	T02531 probable SCARECROW
23	396	12.4	483	2	T10552 SCARECROW-like prote
24	389.5	12.2	583	2	T45597 SCARECROW-like pro
25	384	12.0	573	2	T51239 SCARECROW-like pro
26	344	10.8	410	2	T45848 hypothetical prote
27	304	9.5	623	2	T47874 SCARECROW-like pro
28	298.5	9.3	640	2	B84887 probable SCARECROW
29	293	9.2	486	2	E85433 SCARECROW-like pro

30	288.5	9.0	558	2	T01343 hypothetical prote
31	285.5	8.9	378	2	T51237 SCARECROW-like pro
32	187.5	5.9	205	2	T51233 SCARECROW-like pro
33	179.5	5.6	261	2	T51243 Scil protein (limpo
34	160	5.0	133	2	T51240 SCARECROW-like pro
35	144.5	4.5	1032	2	D83637 serine/threonine p
36	139.5	4.4	1428	2	T13926 probable protein p
37	138	4.3	443	1	T38239 transcription fact
38	134.5	4.2	1151	2	T18535 high molecular mas
39	134	4.2	523	2	T36677 probable secretory
40	133.5	4.2	1420	2	T37781 probable cytoskele
41	132.5	4.1	1487	2	T02850 hypothetical prote
42	130.5	4.1	275	2	T04480 acyl-CoA oxidase h
43	129	4.0	1041	2	T15521 hypothetical prote
44	127.5	4.0	1446	1	A45344 immediate-early pr
45	127	4.0	399	2	I49754 homeobox protein -

## ALIGNMENTS

RESULT	1	2
D84426	hypothetical protein At2g01570 [imported] - Arabidopsis thaliana	
C:Species:	Arabidopsis thaliana (mouse-ear cross)	
C:date:	02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001	
C:Accession:	D84426	
R:lin, X.:	Kaul, S.; Rounsley, S.D.; Shea, T.P.; Bealito, M.I.; Town, C.D.; Fujii, C.Y. M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter Nature 402, 761-768, 1999	
A:title:	Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.	
A:Reference number:	AB4420; M0ID:20083487	
A:Accession:	D84426	
A:Status:	preliminary	
A:Molecule type:	DNA	
A:Residues:	1-587 <STO>	
A:Cross-references:	GB:AE002093; NID:g3785986; PIDN:AAC6733.1; GSPDB:GN00139	
C:geneids:	A:gene: At2g01570	
A:Map position:	2	
Query Match	50.3%;	Score 1608; DB 2; Length 587;
Best Local Similarity	53.6%;	Pred. No. 3.1e-92;
Matches 339; Conservative 82; Mismatches 150; Indels 62; Gaps 13;		
QY	1	MKREYQDAGS---SGGGGCGMGSSDDKMW--SAAAGGEEVDELLAIGYVRRASDMA 54
DB	1	MKRDHHPQGRLSNHGTSSSSSISDKMMVKKEDDGGNMDELLAIGYVRRASSEMA 60
QY	55	DVAQKLEOLEMAMGKGVGAGAPDDSFATHLATDVHYVPTDSSWSVESMLSELNAPP 114
DB	61	EVALKLEQLETFM-----SNVQEDGLSHLATDVHYVNPSELISWLDMLSELN--PP 110
QY	115	PLRPAPQULNASTSYVTGSGGYFDLPSPVD---SSSIVALRPIPSAGATADA-DLSAD 170
DB	111	PLP-----ASSNGLDPLVLPSPICGFPASDYLKYLPCNAIYQFPALDISSSS 157
QY	171	SVADPKRMRTGSGSTS--SSSSSSSLGGARSSVVEAPVAAAANATATLVVVVDITQ 228
DB	158	SNQNRRLKSCSSPDSWVSTGTGTGGVIGTVTTTTTTTAAAGSTBS--VIVDSQ 215
QY	229	ENGIRLVHALLACAEVVOENLSAAEALVQIPLLAASOGAMRKVAAYGEALARRVR 288
DB	216	ENYRVLVHALLACAEALQNNLTLEALVQICLAVSQGAMRKVAATYAEALARIYR 275
QY	289	FRDQPSLLDAAFADLLAHAFYECSPYKFAHFTANQALILEAFACGRVNVVDFGIKQ 348
DB	276	L--SPQNDICHLSDTLQNHFYETCPYKFAHFTANQALILEAFEGKRVNVIDFSNQG 333
QY	349	MORPALLQALALRGGPRPRRLTGVRPORDPDALQOYGVWKLAFHNTIRVDFQVRLV 408

Db 334 LGNRNMAALALREGCGPPTFFRLTGTGPPAPDNDLHIEVGCKLAQALAEIATHEEPEYRFV 393  
 Oy 409 AATLADLEPFLMQPBGEDNEEPEVAVNSVFEMHRLLAQPALEKVLGTVRAVRPRIV 468  
 Db 394 ANSLADLADSLM---ELRPS-DTEAVAVNSVFELHKLILGRPGIEKVLGVAKQIKPIE 448  
 Oy 469 TVVROEAAHNSGTFILDRPTESLHYSTFWDSLEGSSCGGSEVSSGAAAPAAAGTPOV 528  
 Db 449 TVVDESHNHPVFLDRPTESLHYSTFLDSELG-----VPSQDKV 490  
 Oy 529 MSEVYLKQICNVVACBGAERTERHETLGMWRNRLGNAGFETVHLGSNAVYQASTLLALF 588  
 Db 491 MSEVYLKQICNVVACBGPDRVERHERETLSQMGNRFSSGLPAHLGSAWAKQASMLISVF 550  
 Oy 589 AGDGQYKVEEKGCLTGLGMHTRPLIATSMARLA 621  
 Db 551 NSGQGRVYESNCGCLMGHTRPLITTSAMKLS 583  
 RESULT 2  
 H86282  
 protein F10B6\_34 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 R:Accession: H86282  
 R:Ecology: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
 Chn, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.,  
 ansen, N.F.; Hughes, B.; Hutzler, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Lucos, J.S.; Maiti, R.; Marzalli,  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
 Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719  
 A:Accession: H86282  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-533 <STC>  
 A:Cross-references: GB:AE005172; NID:98778219; PIDN:AAF9228.1; GSPDB:GN0141  
 C:Genetics:  
 A:Gene: F10B6\_34  
 A:Map position: 1

[illegible][illegible]

RESULT 3  
G96688  
[hypothetical protein T27F4.10 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: G96688  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chiu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,  
ansen, N.F.; Hughes, B.; Hultzer, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luos, J.S.; Maltl, R.; Marzita  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: G96688  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-511 <STO>  
A:Cross-references: GB:AE005173; NID:g10092507; PIDN:AAG12907.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1  
A:Gene: T27F4.10

Query Match	47.3%	Score 1511.5	DB 2	Length 511
Best Local Similarity	52.3%	Pred. No. 2.5e+86		
Matches 324	Conservative	65	Mismatches 186	Indels 113
				Gaps 13
QY	1 MKRREVQDAGGSGGGGSGGSSSEDKMWSAAGEBVEVELLAALGYKRAADMAVQAOL	60		
	1 MKRRENNHRESSAGEGCG--SSSMITYIKEPAAG---VDLELVVLGYKRRSSMDADVAHKL	54		
Db				
QY	61 EOLEEAMMGVGVGAACAAPDDSPATHLATDTVHYNPTDLSWVESMSELNAPPPLPPL	120		
Db	55 EOLEENVLVDG-----ISNLSDETVIYNRSDLSGWESMLSDLD--PTRIQENP	100		
QY	121 QLNASTSTVTGSGGYFDLPSPVSDSSSITVLRPIPSPAAGATAPADLSADSVYRDKRMFT	180		
Db	101 -----DSEYDLRAI--PGSAVYPRD--EHYVRRSKRTRI	130		
QY	181 GGSSTSSSSSSSSSLGGGARSSVGEAAPVAAANAATPALPVPVVYVDTQAGRLYHALLA	240		
Db	131 ESELSTRS-----VVVLDSEQETGRVHALLA	158		
QY	241 CAEAAYQGNLSAAELVYKQIPELLASOGGAMKRVAYFGCALRVRPRRPDDSLDDA	300		
Db	159 CAEAAYQNNLKLADLVYHVGGLASSQAGAMKRVATYRAGLGARIRYKRPDDVAL--S	216		

[illegible]

	RESULT	4	
	T51475		
	RGA-like protein - Arabidopsis thaliana		
N:	Alternate names:	protein K3M16_50	
C:	Species:	Arabidopsis thaliana (mouse-ear cress)	
C:	Date:	18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000	
C:	Accession:	T51475	
R:	Sato, S.; Nakamura, Y.; Kaneo, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; Mey-		
	submitted to the Protein Sequence Database, August 2000		
A:	Reference number:	Z25394	
A:	Accession:	T51475	
A:	Status:	Preliminary	
A:	Molecule type:	DNA	
A:	Residues:	1-523 <SAT>	
A:	Cross-references:	EMBL:AL391150	
A:	Experimental source:	cultivar Columbia; BAC clone K3M16	
C:	Genetics:		
A:	Map position:	5	
A:	Note:	K3M16_60	
Query Match	45.4%;	Score 1450.5;	DB 2; Length 523;
Best Local Similarity	49.3%;	Pred. No. 1.5e-82;	
Matches 313;	Conservative 76;	Mismatches 115;	Indels 131; Gaps 13;
OY	1 MKRREYQDA-----GGSGGGGGMGSSEDEKMYSAAGBEFEVDLLAALGY 46		
	: :		
Db	1 MKRSHQETSVVEEAPSNWEKLENGCGGG-----DDNM-----DEFLAVLGY 42		
OY	47 KYRASADMAVQAOKLEOLEEMAMGCVGAGAAPPDSSFATHLATDVHYNPTDLSSWVESML 106		
Db	43 KYRSSADMADVAKTEGLEMLV-----SNDLAASSNANENDYHYNPDSLSGAQSML 93		
OY	107 SELNAPPPPLPPAPOLNASTSYTGSGGYFDLPSPVSDSSSIYALRPISPDAGTAPAD 166		
Db	94 SDLVNYP-----DLDP-----NRICDLRPI-----TDDE 118		
OY	167 LSASVDRPKRMRTGGSTSSSSSSSLGGGARSSVVEAPPVAAAMNATPALPVPVVVD 226		
	: : :		
Db	119 CCSSNSNSNKRIIRLGPMCDSTVTSESTRS-----VVL 149		
OY	227 TOEAGIRLVHALLACAFAVOQENESAAEALVKQIPLILASOGAMRKVAAYFGALARRV 286		
Db	150 IETGVRLVQLAVCAEAVALLENISLADALVRVGLLAASQAGANGKVATYFAELARRI 209		
OY	287 FRFRQPQSSILLDAFADLLIAHFIESCPYLKFHAFTLNQAILLEAFACRCRVHVYDFGIK 346		
Db	210 YRI--HPSPAAAIDPSFEEDIIOMNFYDSPYLKFAHTNQAILLEVVTTSRVVHHVIDLEIN 267		

[illegible][illegible]

```

Db 442 ALQATGKRLDFADKLGPFCPL-AEKVGNLDERLVN-----RKREAAVH--WL 491
QY 443 MHRLLAQPGALEKLVGVRAVRPIVVEOEANHNSGTLDRPTESLHYSTPDSLEG 502
Db 492 QHSLYDVYTGSDAHTLMLQRLAPKVYVVEODLSH--AGSFLGRFVEAIIHYSALEPDSL-- 548
QY 503 GSSGCGPSEVSSGAAAPAAAGTDQVNSEVYLGROIQNVVACEGAEERTERHETLGMWRNR 562
Db 549 GASVGESE-----ERHVEEQQLSKEIRNVLAAGSPRSAGEVK--FESNREK 594
QY 563 LGNAGFTVHLGSNAVYQASTLLALFAGDGYKVEKEGCLTIGMHRPLIATSAW 618
Db 595 MQCCGFGISLAGNAATQATLLGMFP--SDGYTLVDNGTLKLGKMDLSLITASAW 649

RESULT 6
T51244
SCARECROW protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 28-Jul-2000 #sequence_revision 28-Jul-2000 #text_change 28-Jul-2000
C:Accession: T51244
R:di Laurenzio, L.; Mysoczka-Dillier, J.; Malamy, J.E.; Pysh, L.; Helariutta, Y.; Freshour
submitted to the EMBL Data Library, July 1996
A:Reference number: Z29338
A:Accession: T51244
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-653 <DIL>
A:Cross-references: EMBL:U62798; PIDN:AA06318.1
C:Genetics:
A:Gene: SCARECROW1
A:Introns: 511/2

```

```

Query Match 18.8%; Score 600.5; DB 2; Length 653;
Best Local Similarity 29.9%; Pred. No. 9.3e-30;
Matches 178; Conservative 84; Mismatches 203; Indels 131; Gaps 18;

```

```

QY 79 DDFATHTADTYVNYNPTDLSWVESMSEL-----NMPPLPAPOLNA 124
Db 129 DQSSSSSSASPTV-----WYDAIIRDLIHSTSVSLPQLIQNVRIITPCNPILGA 179
QY 125 -----STSYTSGSGYFDLPSSVSSSIYALRPIPSGATAPADLSAD 170
Db 180 LLEKRLSLMLDPSSS-----DPSQTFEPLVQISNPSF----- 216
QY 171 SVDRPKMRKRGSGSTSSSSSSSLGGARSSVVEAPPVAAANATPRLPVVVDIOEA 230
Db 217 ----PQQQQHQQQQQHKPPPIQQOERENSTDAPEPETVAT--VPVQTNAAEA 270
QY 231 -----GIRLVHALLACAEAVQENLSAAELVYKQIPLLAASOGAMRKV 274
Db 271 LREKKEIKRKQKODEBGLHLLTLLOCAEAVSADNLEANKLLEISOLSPYTSQKRV 330
QY 275 AAYFGALARRVF-----RFRPOPDSLLDAFAFDLHAHFYESCPLYKFAHF 322
Db 331 AAYFSEMSARLLNSCLGIYAALPSRMMPTHTSLKMSAF-----QYFNGISPLVKFSHF 385
QY 323 TANQALTEARCARVNVDPGKQGMWPAALLQALALRPGGPPSFLTYGPPQPDDET 382
Db 386 TANQALQDEAEKEEDSVHIIIDIMQGLWPGFLHILASRPGPPHVRVLTGLG---TSM 441
QY 383 ALQGVKLAQFAHTIVDQYRGVLAATLADLEPMLQRPGEEDPNEPEVIAVNSVEE 442
Db 442 ALQATGRLSDFTDKLGPFCPL-AEKVGNLDERLVN-----RKREAAVH--WL 491
QY 443 MHRLLAQPGALEKLVGVRAVRPIVVEOEANHNSGTLDRPTESLHYSTPDSLEG 502
Db 492 QHSLYDVYTGSDAHTLMLQRLAPKVYVVEODLSH--AGSFLGRFVEAIIHYSALEPDSL-- 548
QY 503 GSSGCGPSEVSSGAAAPAAAGTDQVNSEVYLGROIQNVVACEGAEERTERHETLGMWRNR 562

```

```

Db 549 GASVGESE-----ERHVEEQQLSKEIRNVLAAGSPRSAGEVK--FESNREK 594
QY 563 LGNAGFTVHLGSNAVYQASTLLALFAGDGYKVEKEGCLTIGMHRPLIATSAW 618
Db 595 MQCCGFGISLAGNAATQATLLGMFP--SDGYTLVDNGTLKLGKMDLSLITASAW 649

```

```

RESULT 7
E96540
hypothetical protein FILF12.22 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: E96540
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Hultzar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Matil, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: E96540
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-482 <STO>
A:Cross-references: GB:AE005173; NID:99454566; PIDN:AAF87889.1; GSPDB:GN00141
C:Genetics:
A:Gene: FILF12.22
A:Map position: 1

```

```

Query Match 17.9%; Score 572.5; DB 2; Length 482;
Best Local Similarity 28.7%; Pred. No. 3.4e-28;
Matches 147; Conservative 79; Mismatches 171; Indels 115; Gaps 10;

```

```

QY 181 GGSSTSSSSSSSLGGARSSVVEAPPVAAANATPRLPVVVDIOEAGIRLVHALLA 240
Db 10 GTSSVASPPLOVFSTMSLNRPRTLLASSPFHCLKDLP-----EERLYLHLLLT 60
QY 241 CAEAVQENLSAADLVYKQIPLLAASOGAMRKVAAVFGALARRVFRFPQDSSLDA 300
Db 61 CANHVASGSIONANAALEQSHLASPDGDTMORIAAVFTEALANRLIKSWP----GLYKA 116
QY 301 AFA-----DILHMH--FYESCPLYKFAHTANQALILEAFAGCRRVNVVDGKQGM 349
Db 117 LNAIQTRTNNVSEIHYRRLFEEMFPILKYSLTLNRLALEMEGKKVAVHVIDDASEPA 176
QY 350 QWPAALLQALALRPGGPPSFLTYGPPQPDDETALQGVKLAQFAHTIRVDFQYGLVA 409
Db 177 QWMLALQAFNSRPGPPHRLRTGV-----HQKEVLEQMAHRLIEAEKLDLPQGFNPVYS 232
QY 410 ATLADLEPMLQRPGEEDPNEPEVIAVNSVEMHRLA----- 448
Db 233 RLDCIANEQLRV-----KTGEALAVSVSLQHLTFIASDDDLMKKNCALRFQNNPSG 283
QY 449 -----QPGALEKLVGVRAVRPRI 467
Db 284 VDLQRLVIMSHGSAARENDMSNNNGSPSGSASSLPPLPSSGRTSFLMAINGLSKV 343
QY 468 VTYVEOEANHNSGTLDRPTESLHYSTPDSLEGSSGGSGPSEVSSGAAAPAAAGTDQ 527
Db 344 MVYTEQDSDNHSGTLMERLLESLTYAALPDCLE-----TYVPRTSQDRI 388
QY 528 VMSEVYLGROIQNVVACEGAEERTERHETLGMWRNLGNAGEVTVHLGSNAVYQASTLLAL 587
Db 389 KVEKMLDGEELKNITISCEGFERRHRHELEKWSQRIDLAGGVNPLSYAMLAQARILLQ- 447
QY 588 FAGDGYKVEKEGCLTIGMHRPLIATSAW 619
Db 448 GCGFDGYRIKEESCAVITCWQDRPLYSAMR 479

```







[illegible]

A:Reference number: Z25337; MUID:99272994  
A:Accession: T51234  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-352 <PYS>  
A:Cross-references: EMBL:AF036300; PIDN:AAD24403.1  
C:Genetics:  
A:Gene: SCUL

Query Match                      14.5%; Score 464; DB 2; Length 352;  
Best Local Similarity        31.4%; Pred. No. 1,2e-21;  
Matches    120; Conservative    67; Mismatches 147; Indels    48; Gaps    10;

QY    255 ALVKQIPLLAASGGGAMKVAAYFGALARRV----FRFR---POPDSSLLDAFAFL 305  
         :::::  
Db     2 SMVNELRQIVSIQGDPDSQRIAAYWEGIAABMAASGFIFYALKCKEPPSDERLAQVL 61  
         :::::  
QY    306 LHAHFESCPFLKRAHNFANQAIIIEFAGCGRRVAVDVGITQGNQWALLDALLRGCP 365  
         :::::  
Db     62 -----FEVCPCPFKGFFLAANGAILLEAIKGEBEVHIIDFDINQNGQMYTLIRLSIELPGKR 116  
         :::::  
QY    366 PSFRLTGVGPPOPESTD--ALOQYGWKLAOPAHNTTRVDFOYRGVAATLADLERFML~-Q 421  
         :::::  
Db    117 PRLRLTGIDDEPSVORSIGGLRIITGLRLEQLAEONGVSFKRKAPSKT-SIVSSTLGCK 175  
         :::::  
QY    422 PEGEDPNEEPEEVIAVNSVFEMHL---LAQGALEKVLGTBAVRPRITYVEQEANH 477  
         :::::  
Db    176 P-----SETLLVPFAFOLHHMPDESVTTVNQRDPELLHMVSLPKLTVVVEODVNT 226  
         :::::  
QY    478 NSGFTLDFTESLNHYTSIMPDSLGGSGGSPSEVSSGAAAPPAAGIDQVMSEVIYGRQ 537  
         :::::  
Db    227 NTSPFPFRIFAYEYSASFESLD-----MTLPRESQERNMVERQCILD 271  
         :::::  
QY    538 ICNVACGAERTERHEHTLGGMRNLGNAGEFYHLGSNAVKQASTLLTPAGDGXYKE 597  
         :::::  
Db    272 IVNIYACGGERIERIEYEAAGKKRARMMAGENPRPMASKVTNNIQLIKQ-QYCNKYTKL 330  
         :::::  
QY    598 EEKGLTLGMWTRPLIATSANR 619  
         :::::  
Db    331 EEMGELHFCWEKSLIVASAWR 352  
         ::::: :

RESULT    15  
T02736  
probable SCARECRON gene regulator [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 02-Feb-2001  
C:Accession: T02736; A84692  
R:Rounsley, S.D.; Lin, X.; Kaul, S.; Shea, T.P.; Fujii, C.Y.; Mason, T.M.; Shen, M.,  
submitted to the EMBL Data Library, August 1998  
A:Description: Arabidopsis thaliana chromosome II BAC T914 genomic sequence.  
A:Reference number: Z14710  
A:Accession: T02736  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1336 <ROU>  
A:Cross-references: EMBL:AC005315; NID:g3461834; PID:g3461846  
A:Experimental source: cultivar Columbia  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.  
M.; Koo, H.; Moffit, K.S.; Cronin, L.A.; Shen, M.; Vanhaken, S.E.; Umayam, L.; Tallon,  
cuss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
Nature 402, 761-768, 1999  
A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487  
A:Accession: A84692  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1336 <STO>  
A:Cross-references: GB:AE002093; NID:g3461846; PIDN:AAC33232.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: AT2g29060; T914.14  
A:Map position: 2



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2002, 16:24:50 ; Search time 17.09 Seconds  
(without alignments)  
1336.582 Million cell updates/sec

Title: US-09-485-529-7  
Sequence: 1 MKREYDAGSGGGGGGSGGMS.....TLGHTPLATLSAWRLAGP 623

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues  
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SWISSProt\_39:\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	152	4.8	375	1	SOX3_MOUSE
2	138	4.3	443	1	SOX3_HUMAN
3	136.5	4.3	1461	1	IE18_PRIVF
4	136	4.3	762	1	P115_CHICK
5	130.5	4.1	393	1	HXAA_HUMAN
6	130.5	4.1	507	1	IRX3_MOUSE
7	127.5	4.0	1446	1	IE18_PRIVK
8	127	4.0	399	1	HXAA_MOUSE
9	124.5	3.9	681	1	YDHE_SCHPO
10	124.5	3.9	825	1	ICP0_HSV2H
11	123.5	3.9	1676	1	APSA_EMENT
12	123.5	3.9	3038	1	TRIO_HUMAN
13	123	3.8	606	1	CENB_CRIGR
14	122.5	3.8	901	1	ALB0_MOUSE
15	119.5	3.7	2493	1	CYAA_USTMA
16	118.5	3.7	3567	1	ERY2_SACER
17	118	3.7	483	1	PREG_NEUCR
18	118	3.7	975	1	CUT1_CANFA
19	118	3.7	1395	1	CUT1_MOUSE
20	117.5	3.7	474	1	SOX4_HUMAN
21	117.5	3.7	915	1	ALB0_RAT
22	117.5	3.7	1059	1	CAPU_DROME
23	116.5	3.6	1001	1	ORL1_DROME
24	116.5	3.6	1093	1	AF17_HUMAN
25	116.5	3.6	1487	1	ICP4_HSVK
26	116	3.6	870	1	BCAL_HUMAN
27	116	3.6	889	1	CIKE_RAT
28	116	3.6	1294	1	CUT2_HUMAN
29	115.5	3.6	728	1	VIV_ORYSA
30	115	3.6	891	1	MAZ3_SCHCO
31	115	3.6	2442	1	CBP_HUMAN
32	114.5	3.6	1487	1	ICP4_HSVB
33	114.5	3.6	3421	1	TBCU_HSVB

34	114	3.6	500	1	BRN1_HUMAN
35	114	3.6	721	1	YR82_MYCTU
36	114	3.6	746	1	7UP2_DROME
37	114	3.6	1229	1	NI21_HUMAN
38	114	3.6	1298	1	ICP4_HSV1
39	113.5	3.6	1322	1	YAG3_YEAST
40	113.5	3.6	2038	1	F5H_DROME
41	113	3.5	995	1	Y109_YEAST
42	113	3.5	1367	1	AMTH_YEAST
43	113	3.5	2090	1	N214_HUMAN
44	113	3.5	2441	1	CBP_MOUSE
45	113	3.5	5179	1	MUC2_HUMAN

## ALIGNMENTS

RESULT 1	ID	SOX3_MOUSE	STANDARD	PRT	375 AA.
AC	P53784				
DT	01-OCT-1996 (Rel. 34, Created)				
DT	01-OCT-1996 (Rel. 34, Last sequence update)				
DT	15-JUL-1998 (Rel. 36, Last annotation update)				
DE	TRANSCRIPTION FACTOR SOX-3.				
GN	SOX3 OR SOX-3.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_Taxid=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=129;				
RX	MEDLINE=96189340; PubMed=8625802;				
RA	Collignon J., Sockanathan S., Hacker A., Cohen-Tannoudji M.,				
RA	Norris D., Rastan S., Stevanovic M., Goodfellow P.N.,				
RA	Lovell-Badge R.;				
RT	* A comparison of the properties of Sox-3 with Sry and two related				
RT	genes, Sox-1 and Sox-2.*				
RL	Development 122:509-520(1996).				
CC	-1- SUBCELLULAR LOCATION: NUCLEAR.				
CC	-1- TISSUE SPECIFICITY: MAINLY IN THE DEVELOPING CENTRAL NERVOUS				
CC	SYSTEM. EXPRESSED IN DEVELOPING UROGENITAL RIDGE.				
CC	-1- SIMILARITY: CONTAINS 1 HMG BOX.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	-----				
DR	EMBL: X94125; CAA63845.1; -				
DR	HSSP: Q05066; 1HR2.				
DR	MGD: MGI:98365; Sox3.				
DR	InterPro: IPR000910; HMG_12_box.				
DR	Pfam: PF00505; HMG_box; 1.				
DR	SMART: SM00398; HMG; 1.				
KM	DNA-binding; Nuclear protein; Transcription regulation.				
FT	DOMAIN 54				
FT	DNA_BIND 63				
FT	FT 69				
FT	DOMAIN 164				
FT	DOMAIN 177				
FT	DOMAIN 219				
FT	DOMAIN 223				
FT	DOMAIN 252				
FT	DOMAIN 259				
FT	DOMAIN 276				
FT	DOMAIN 282				
FT	DOMAIN 293				
FT	SEQUENCE 375 AA; 37857 MW; 2A8477A10A517EEF CRC64;				

Query Match 4.8%; Score 152; DB 1; Length 375;  
Best Local Similarity 20.4%; Prod. No. 0.023;  
Matches 69; Conservative 36; Mismatches 133; Indels 100; Gaps 8;

Oy	8	AGSGGGGCGGCSSEED-----KIMVSAAGEGVDDLLA-----	43
Dd	52	ANGGGGGGGGGQDVKRRPMNFWMYMSGQRKKMALLEPKHNISIRLGADMKLT	111
Oy	44	-----LCYKRASDMADVQCKLEQLLEAMMGVGGAARDDS	81
Dd	112	DAEKRPTEAKRLRAVMHEKPYKKRPRRKTKTLKKDKKYSLPGGLPPRGAAAAAAA	171
Oy	82	FATHLADFTVHYNP-----FDLSWVESMLT-----ELNAPPP-----	115
Dd	172	AAAAAAAASSPVGVGRDLDTYTHVNNGANGAVSYLGEOLQAOPRSMSPPPRALPOMHRY	231
Oy	116	-----LPRAPLNASTNSSTVTWGSGCFDLPRS-VDSSTSYALRLPIPSAGTAR	164
Dd	232	DMAGSLQYSMPMRPAOASYMNAAAAAAASAAGGCAFPAAAAAAAAYOOQRYTAIAAAAA	291
Oy	165	ADLSADSVRDPKRMKRTGGSSPTSSSSSSSSSSSGLGGARRSSVEEAPRVAAANATPALRVVV	224
Dd	292	AAMSIGPWGSVVKSEPSPPRALISHQORACLDGLRLMISMUTLRPPGDADAADSPLP---	348
Oy	225	VDTQEGRIYLVALLACAENAOGENLSAAETLYKOIPL	262
Dd	349	-----GGRL-----HGYNHQYGAGTYNGTVNCSTVR	372

RESULT	2			
SOX3_HUMAN				
ID	SOX3_HUMAN	STANDARD:	PRF:	443 AA.
AC	P41225; P35714;			
DT	01-JUN-1994 (Rel. 29, Created)			
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DT	15-JUL-1998 (Rel. 36, Last annotation update)			
DE	TRANSCRIPTION FACTOR SOX-3.			
GN	SOX3.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RP	MEDLINE=94154672; PubMed=8111369;			
RA	Stevanovic M., Lovell-Badge R., Collignon J., Goodfellow P.N.;			
RT	"SOX3 is an x-linked gene related to SRX.";			
RL	Hum. Mol. Genet. 2:2013-2018(1993).			
RP	[2]			
RP	SEQUENCE OF 150-203 FROM N.A.			
RX	MEDLINE=92310993; PubMed=1614875;			
RA	Denny P., Swift S., Brand N., Dabhadre N., Barton P., Ashworth A.;			
RT	"A conserved family of genes related to the testis determining gene,			
RL	SRX.";			
RL	Nucleic Acids Res. 20:2887-2887(1992).			
CC	-1- SUBCELLULAR LOCATION: NUCLEAR.			
CC	-1- SIMILARITY: CONTAINS 1 HMG BOX.			
CC	-1- CAUTION: WAS CALLED SOX-9 BY REF. 2.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL: X71135; CAA50465.1; -;			
DR	EMBL: X65665; CAA46616.1; -;			
DR	PIR: S21483; S21483.			
DR	PIR: S22942; S22942.			
DR	HSSP: 005066; 1HRZ.			
DR	MIM: 313430; -;			
DR	InterPro: IPR000910; HMG_12_box.			
DR	Pfam: PF00305; HMG_box; 1.			
DR	SMART: SM00398; HMG; 1.			

Protein	Length (aa)	Accession	Function
KM DNA-binding; Nuclear protein; Transcription regulation	129	133	POLY-GLY.
FT DOMAIN	139	207	HMG BOX.
FT DNA_BIND	139	248	POLY-ALA.
FT DOMAIN	234	268	POLY-PRO.
FT DOMAIN	290	294	POLY-ALA.
FT DOMAIN	321	327	POLY-ALA.
FT DOMAIN	337	344	POLY-ALA.
FT DOMAIN	350	361	POLY-ALA.
FT DOMAIN	159	159	L -> Q (IN REF. 2).
FT CONFLICT	159	159	D -> E (IN REF. 2).
FT CONFLICT	176	176	E -> D (IN REF. 2).
FT CONFLICT	202	202	
EQ SEQUENCE	443 AA;	44884 MM;	8031B4EDAD52DB34 CRC64;

Query Match	4.38;	Score 138;	DB 1;	Length 443;
Best Local Similarity	20.28;	Pred. No. 0.19;		
Matches	69;	Conservative	37;	Mismatches 127;
				Indels 108;
				Gaps 10.

Qy	9	GGSGGG--GGGMSSEB-----KMWVAAGGEGVEVELLAA-----43
Db	121	GGSSGGAGGGGGGTDQDVRKRMNAFWMSRGQRKKMALENPKMHNSSEISKRGADWKLTL180
Qy	44	-----LGYVRASDADVAQKLEOLEEMAMGCGVGAGAADPD80
Db	181	TDAEKRPFIDEAKRLRAVHNMEKYRDPDYRYRPRKRTKTLKKDKTISLPSGLLPPGAAAAAAA240
Qy	81	SEATHLAT-----DTVHNPTDLSSWVESMLS-----ELNAPPP--115
Db	241	AAAAAAAAAASPVGQGRLDYF---TVHNGMANGATYSLVQEQLEAYDAPPMSSSPPPPPAL296
Qy	116	-----LPPAPQLNASTSSVTYGGGCTFDLPSPVD--SSSIYALRLPRLPSAGA161
Db	297	HRYDWAGLQYSPMPMPGAQSYMNVAIAAAAAASGGGMAPSPSTAAAAAAYTQOQPTAAAA356
Qy	162	TAPADLSADSVRDKRMRTGGSTSSSSSSSSSILGGGARSSVVEADPVAANAANATPAL221
Db	357	AAAAAMSLGPRGSSVYKSPSSPPPALINASHQRACTGLRLDMISMYLRPPGDADADAASPLP416
Qy	222	VVVVDTQEAIGIRLVHALLACAEAVQOENLSAAELVYQICPL262
Db	417	-----GGRL-----HGVAHQHVGAGTAVNGCTPL440

```

RESULT 3
ID IE18_PRIVT STANDARD; PRT; 1461 AA.
AC IE18_PRIVT
PI 1675;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-APR-1990 (Rel. 14, last sequence update)
DT 01-FEB-1994 (Rel. 28, last annotation update)
DE IMMEDIATE-EARLY PROTEIN IE180.
GN IE.
OS Pseudorabies virus (strain Indiana-Funkhauser / Becker) (PRV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicelloviruses.
OX NCBI_TaxID=31523;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89315207; PubMed=2546124;
RA Cheng A.K.;
RT "DNA nucleotide sequence analysis of the immediate-early gene of
RL Pseudocytic virus."
RN Nucleic Acids Res. 17:4637-4646(1989).
RN [2]
RN REVISIONS.
RA Cheng A.K.;
RL Submitted (NOV-1989) to the EMBL/GenBank/DBJ databases.
CC -! FUNCTION: THIS IE PROTEIN IS A MULTIFUNCTIONAL PROTEIN CAPABLE
CC OF MIGRATING TO THE NUCLEUS, BINDING TO DNA, TRANS-ACTIVATING
CC OTHER VIRAL GENES, AND AUTOREGULATING ITS OWN SYNTHESIS.
CC -! SUBCELLULAR LOCATION: NUCLEUS OF INFECTED CELLS.
CC -! PRT: A LONG STRETCH OF SERINE RESIDUES MAY BE A MAJOR SITE OF
CC PHOSPHORYLATION.

```

CC -1- SIMILARITY: BELONGS TO THE ICP4/IE140/IE180 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: X15120; CAA33214.1; -  
CC PIR: S04713; EDBE1F.  
CC Early protein; Transcription regulation; Trans-acting factor;  
CC DNA-binding; Phosphorylation; Nuclear protein.  
CC DOMAIN 390 405 POLY-SER.  
CC FT DOMAIN 958 966 POLY-SER.  
CC SEQUENCE 1461 AA; 149833 MW; 7F31E7ABE403B208 CRC64;

Query Match 4.3%; Score 136.5; DB 1; Length 1461;  
Best Local Similarity 22.1%; Pred No. 1;  
Matches 158; Conservative 52; Mismatches 255; Indels 251; Gaps 30;

OY 8 AGSGGGGGGGSSSEDDKMYNSAAAGE-GE-----EYDELLAALGYEVNAS 51  
DB 179 SSGSSSSSSSSPSPAPARRMSPARGDPEGPAPARRTPAPAPQAPAAVAANARRGPAS 238  
OY 52 DNADVAQKLEQLAMAMGCGVGAAPDDSFATHLATDTVHYNPDTLSSWESMSELNA 111  
DB 239 PASPAAGPVY-----APGCGGAPSGGDCGRHH-----HQHR-----EPLDEPAA 279  
OY 112 -----PPPL-----PPAPQANASTSTVY-----GSGGFDLP----- 140  
DB 280 ARRLDPRPLGANSVPSNPNSSSTTTVAVEPVANGPEKEDDGLAGDGG-----APLOR 336  
OY 141 -----PSVDSSTSYALRPIPSAGATAPADLSADSVDRPKRMRTG 182  
DB 337 QRRRRRAGGALRGNGFGSSSSGSDSLSPARSPAPRAAAAAA-----RRASSS 392  
OY 183 SSTSSSSSSSSSLGGGANSVVEAPPVAAANATPALPVVVVDTGEGIRLVHALLAC 242  
DB 393 SSSSSSSSSSSSEGEDEGCVRGAPLARAGPPPPAPAPAPASASATSSSAAS 452  
OY 243 EAVQOE-----NLISAALVKGQILLA-----ASOGGA 270  
DB 453 PAPAPAPAPRRKRSTNNHUSLMADGPPPTDGLTLPLGEFWGSDPPADGRVRYGA 512  
OY 271 -----MRKYAAYFGEALARVFRFRPO-PDSSLDAAFADLHANFY-ESCPY 316  
DB 513 GDSREGIMDEDDYROAAARY-RAAGRPVVFIPEDGDSKQNHALVRLIYSGAAGFAMSW 571  
OY 317 LKFAHTANQAILLEAFAGC-RRYHVVDVFIGKQGMOPALLQALALRPGGPPSFRLTGVP 375  
DB 572 LONPRMQADDOREFNCF--CORRYHA-----PHGHSFTTGSVTP 608  
OY 376 POPDEIDAL--QQVGKKLQAFHTIRV-----DFQYGLVA--- 409  
DB 609 PLPHIGDAADADPLMALPHAVASAMSRRYDRTOKFTLQSLRRAYADMAVYGRADPR 668  
OY 410 -----ATLADLEPMLQPEGEDDEPNEBE-----VIAVNSYFEMHRLI--- 447  
DB 669 AGATATVEALCARVRAFAAAGRGVPRRELADACVLAQCVLE--RLPCLRLPAPARAP 726  
OY 448 --AOGCALKEKVLGVAVRPRI--VTVEQEAHNSGTLDRFTESLHYIS----- 494  
DB 727 AALGPALCEEVYAAALALDAIPGAGPAPROQADSAVALVARTVAVLVSYVSGARARE 786  
OY 495 -----TMF-----DSLEGGSSGGSPSEVSSG-----AAAAAPAG 524  
DB 787 AMTYAALAFAPANVAGARLAEAARPGAPAPGDLPLPMPDEPGVLVAPAPAPAG 842

P115.CHICK  
ID P115.CHICK STANDARD: PRT; 762 AA.  
AC 098917;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE 115 KDA MELANOSOMAL MATRIX PROTEIN PRECURSOR.  
GN MPM115.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-WHITE LEGHORN; TISSUE-Retinal pigment epithelium;  
RX MEDLINE=92020667; PubMed=1924173;  
RA Mochli M., Agata K., Kobayashi H., Yamamoto T.S., Eguchi G.;  
RT "Complete sequence and expression of a cDNA encoding a chicken  
RT 115-kDa melanosomal matrix protein.";  
RT pigment Cell Res. 4:41-47(1991).  
RL [2]  
RN CHARACTERIZATION.  
RP STRAIN-WHITE LEGHORN; TISSUE-Retinal pigment epithelium;  
RX MEDLINE=88311098; PubMed=3409326;  
RA Mochli M., Agata K., Kobayashi H., Yamamoto T.S., Eguchi G.;  
RT "Expression of gene coding for a melanosomal matrix protein  
RT transcriptionally regulated in the transdifferentiation of chick  
RT embryo pigmented epithelial cells.";  
RL Cell Differ. 24:67-74(1988).  
CC -1- FUNCTION: MIGHT BE REQUIRED FOR POLYMERIZATION OF MELANIN ONTO THE  
CC CORE STRUCTURE OF MELANOSOMES WITH ENZYMIC FUNCTION OF TYROSINASE.  
CC -1- SUBCELLULAR LOCATION: ON THE FIBROUS MATRIX STRUCTURE OF THE  
CC PREMELANOSOME.  
CC -1- TISSUE SPECIFICITY: SPECIFIC TO PIGMENTED EPITHELIAL CELLS AND  
CC MELANOCYTES. NOT EXPRESSED IN LENS, NEURAL RETINA, BRAIN, HEART,  
CC GIZZARD OR LIVER.  
CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING THE REDIFFERENTIATION OF  
CC PIGMENTED EPITHELIAL CELLS (PEC).  
CC -1- PTM: GLYCOSYLATED.  
CC -1- SIMILARITY: BELONGS TO THE PMEL-17/NNB FAMILY.  
CC -1- SIMILARITY: CONTAINS 1 PKD DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: D88348; BAA13589.1; -  
CC InterPro: IPR000601; PKD\_domain.  
CC Pfam: PFO0801; PKD. 1.  
CC SMART: SM00089; PKD. 1.  
CC PROSITE: PSS0093; PKD. 1.  
KW Signal; Glycoprotein; Repeat.  
FT SIGNAL 1 19  
FT CHAIN 20 762  
FT DOMAIN 223 323  
FT DOMAIN 441 532  
FT REPEAT 441 464  
FT REPEAT 465 488  
FT REPEAT 489 508  
FT REPEAT 509 532  
FT CARBOHYD 111 111  
FT CARBOHYD 115 115  
FT CARBOHYD 346 346  
FT CARBOHYD 651 651  
FT CARBOHYD 659 659  
SO SEQUENCE 762 AA; 77356 MW; 172C8DB4DFCE7C6 CRC64;









DR SMART: SM00389: HOX.1.  
 DR PROSITE: PS00027: HOMEBOX\_1: 1.  
 DR PROSITE: PS0071: HOMEBOX\_2: 1.  
 KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
 KW Transcription regulation; Alternative splicing.  
 FT DOMAIN 26 38 GLY-RICH.  
 FT DOMAIN 77 90 POLY-GLY.  
 FT DOMAIN 113 144 GIN/PRO-RICH.  
 FT DOMAIN 215 222 POLY-GLY.  
 FT DOMAIN 271 274 POLY-GLY.  
 FT DNA\_BIND 325 384 HOMEBOX.  
 FT VARSPIC 1 305 MISSING (IN ISOFORM 2).  
 FT VARSPIC 306 308 DSL-> MOQ (IN ISOFORM 2).  
 SQ SEQUENCE 399 AA: 41415 MW: 7529624FC6057042 CRC64;

Query Match 4.0%; Score 127; DB 1; Length 399;  
 Best Local Similarity 22.1%; Pred. No. 0.76;

Matches 88; Conservative 37; Mismatches 157; Indels 116; Gaps 18;

QY 9 GSGGGGGGMSSEDKMY/SAAAG--EGEEVDLLAALGYKVRASMDVYQKLEOLEMA 66  
 ||| ||| : : : : :  
 DB 29 GSGAGGGGGGYAHGGVYLPASDLPYGLQSCGLFPALG-----SKRNEAPSP 76  
 ||| ||| : : : : :  
 QY 67 MGVGVGAGAAPDPSFATLATDTVHNPDLSSWESMLS----- 107  
 ||| ||| : : : : :  
 DB 77 GGGGGGGSG-----LGGTGTAPAPLDLWDAPRSCMEPPDPPOPOPOO 127  
 ||| ||| : : : : :  
 QY 108 -ELNAPPELPAPQLNASTS---TVTSGGY--FDLPVSUSSSIVALRP---PSP 158  
 ||| ||| : : : : :  
 DB 128 QQQPPPPPPQPPQPPQANSCSFAONIKESVCLVDADKCPKSAADLAFPPGPP 167  
 ||| ||| : : : : :  
 QY 159 AGTATADLADSVRPPKRTGGSSSSSSSSSLGGAR--SSVEAPVAAAANA 216  
 ||| ||| : : : : :  
 DB 188 DGGALCA---SSGVPPGYFRL--SOAYGAKFGSGGGTEOLASFPAPQP--GRGFD 241  
 ||| ||| : : : : :  
 QY 217 TPALPVVVDTOGATRLVHALLACAEVQENLSAAELVKQIP---LLAASOGAMRK 273  
 ||| ||| : : : : :  
 DB 242 PPLP--ASGSTEAAG-----KERVLDSTPPPTLVCTGGGSGQD 278  
 ||| ||| : : : : :  
 QY 274 VAAVFEALARRVRRPPQDSSILDAAPADLLAHFYES-----CPYLK- 318  
 ||| ||| : : : : :  
 DB 279 EEAHASSAAEEL--SPASENSKASPEKDSGSGKGEAANAWMLAKSKKKRCPTTKH 335  
 ||| ||| : : : : :  
 QY 319 -----FAHFTANQALIEAFACGRVHVYDEGIK 346  
 ||| ||| : : : : :  
 DB 336 QTLLEKEFLFNMYLTRERLEI---SRVHLNDROVK 370  
 ||| ||| : : : : :

RESULT 9

ID YDHE\_SCHPO STANDARD; PRT; 681 AA.  
 AC 092359;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE HYPOTHETICAL 73.3 KDA PROTEIN C6G9.14 IN CHROMOSOME I.  
 GN SPAC6G9.14.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 NC NCBL\_TaxID=4896;  
 RN NCBL\_TaxID=4896;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RA Murphy L., Harris D., Bartell B.G., Rajandream M.A., Connor R.E.;  
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBI databases.  
 CC -1- SIMILARITY: BELONGS TO THE PUMILIO/MP5 FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: 281317; CAB03616.1; -  
 DR InterPro: IPR001313; PUM.  
 DR Pfam: PF00806; PUF; 8.  
 DR SMART: SM00025; Pumilio; 8.  
 KW Hypothetical protein.  
 SQ SEQUENCE 681 AA: 73292 MW: 755A25C68A92D727 CRC64;

Query Match 3.9%; Score 124.5; DB 1; Length 681;  
 Best Local Similarity 20.2%; Pred. No. 2.1;

Matches 95; Conservative 74; Mismatches 135; Indels 167; Gaps 23;

QY 81 SFATLATDTVHNPDLSSWESMLSSELNAPPELPAPQLNASTSVTVGSGYFDLP 140  
 ||| ||| : : : : :  
 DB 113 SFTTSHSSSTTFPPATMGG-----LNLSSP-----SSLYLPGSASY-QRS 155  
 ||| ||| : : : : :  
 QY 141 PSYDSSSIY---ALRPD-----SPGATAPADLSAD-----VRDPK 176  
 ||| ||| : : : : :  
 DB 156 NSKNSSASILQMTLDDIPILRRPQLSSYTPGPSTSRSSSSSLGNGPLIANNPS 215  
 ||| ||| : : : : :  
 QY 177 RMR-----TGGSSSSSSSSSLGGARS--SYVEAPVAAA-----NATPALPVY 224  
 ||| ||| : : : : :  
 DB 216 ASKNFAFTSSSSSTNNSTSSSMANGLQSIKSAFAFLPSLNTSFFGENTLPISLA 275  
 ||| ||| : : : : :  
 QY 225 V--DIOAGIRLVHALLACAEVQOE-----NLAAE-----ALVQIPLAASOG 268  
 ||| ||| : : : : :  
 DB 276 ASGTSDSSSNANTLIATPTLTPSPANTVGNPSADTPGNNVPELIDDDSVSSLS 335  
 ||| ||| : : : : :  
 QY 269 GAMRKVAAFYGEALA-----RVFRFRPPQDSSILDAAPADLLAHFYESCPYLKFA 320  
 ||| ||| : : : : :  
 DB 336 SVAASLSLQNSNLSIFCKDHGCRYLQRLLEKKNQSHIDAVFA-----ETHPYLAV- 386  
 ||| ||| : : : : :  
 QY 321 HFTANQALIEAFAG--CRRVHVVDFTGKQGMQWPAALLQALALRPGPPRLTGVGPP 378  
 ||| ||| : : : : :  
 DB 387 -----LMVDAFGNYLCOKL---FEHASQNSTFIQITAPKI--VPSGNMG----- 429  
 ||| ||| : : : : :  
 QY 379 DETDALQOVGMKLAQFAHTTRVDFOYRGVLAATLADLEPMLQPEGEDENEBEVIAVN 438  
 ||| ||| : : : : :  
 DB 430 --TRALQKI-----ID----- 438  
 ||| ||| : : : : :  
 QY 439 SVEEMRLAQPALKEKVLGTVAVRPRIVYVEQ--EANNHSGTELDRETE 488  
 ||| ||| : : : : :  
 DB 439 -----LVSSP---DOISCIYNALRPVNVILTKDLNGNHVIOKLNKFSQ 479  
 ||| ||| : : : : :

RESULT 10

ID ICP0\_HSV2H STANDARD; PRT; 825 AA.  
 AC P28284;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE TRANS-ACTING TRANSCRIPTIONAL PROTEIN ICP0 (VM118 PROTEIN).  
 GN RL2.  
 OS Herpes simplex virus (type 2 / strain HG52).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Simplexvirus.  
 NC NCBL\_TaxID=10315;  
 RN NCBL\_TaxID=10315;  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9213549; PubMed=1662697;  
 RA McGeoch D.J., Cunningham C., McIntyre G., Dolan A.;  
 RT "Comparative sequence analysis of the long repeat regions and  
 RT adjoining parts of the long unique regions in the genomes of herpes  
 RT simplex viruses types 1 and 2."  
 RL J. Gen. Virol. 72:3057-3075(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.

RA Dolan A.;  
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.  
CC -1- SIMILARITY: TO OTHER HERPESVIRUSES ICPO PROTEIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D10471; BAA23427.1; -  
DR EMBL; 286099; CAB06760.1; -  
DR PIR; J01501; EDBEXD.  
DR InterPro; IPR001841; Znf\_fing.  
DR Pfam; PF00097; zf-C3HC4; 1.  
DR SMART; SM00184; RING; 1.  
DR PROSITE; PS00518; ZINC\_FINGER\_C3HC4; 1.  
KW Transcription regulation; Trans-acting factor; Activator; Zinc-finger;  
KW DNA-binding.  
FT DOMAIN 120 123 POLY-GLY.  
FT ZN\_FING 126 152 RING-TYPE.  
FT DOMAIN 126 152 POLY-GLY.  
FT DOMAIN 266 271 POLY-GLY.  
FT DOMAIN 292 295 POLY-SER.  
FT DOMAIN 342 345 POLY-ALA.  
FT DOMAIN 386 389 POLY-SER.  
FT DOMAIN 395 400 POLY-GLY.  
FT DOMAIN 425 428 POLY-ALA.  
FT DOMAIN 590 627 POLY-SER.  
SQ SEQUENCE 825 AA; 81986 MW; SCEBI585853A274 CRC64;  
  
Query Match 3.9%; Score 124.5; DB 1; Length 825;  
Best Local Similarity 25.9%; Pred. No. 2.7;  
Matches 67; Conservative 22; Mismatches 95; Indels 75; Gaps 10;  
  
OY 11 SGGGGGCGSSGSDKM-----MVSAAAGEGEVDELLALGYKVR 49  
DB 394 SGPGGGGLPQSSGRRAARPAVPRSPRAAAPVVSASADAAGPAPPAVPAHRRP 453  
OY 50 ASDMADV-----AQRLEOLEM--AMGCGVGA-----GAAPDDSFATHLATD 89  
DB 454 RSRMTQACTDTCQASLGRAGATDARGSGPGEAGPGVPRGTNTGAAPH-----AAE 506  
OY 90 TVHYNP-----TDLSSWESMSELNAPPPPLP-----APQLNASTSTVTGSGCY 136  
DB 507 GAAPRRRRRGSDGSPAASSSSASAAAPRSPPLAPQGVGAKRAAPRRAPDSDS---GDRGH 563  
OY 137 FDL-----PPVSDSSSIYALRPIPPAGATAPADISADSVDPKRMRTGSGSTSS 188  
DB 564 GPLAASAGAAPSPSPSQAAVVAASSSSASSSSSSSSSSSSSSSSSSSSSSSSSSS 620  
OY 189 SSSSSSLGGGARSSVEAA 207  
DB 621 SSSASSAGAGGSVAASAS 639  
  
RESULT 11  
APSA\_EMENT STANDARD; PRT; 1676 AA.  
ID APSA\_EMENT  
AC Q00083;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE ANOCLATE PRIMAARY STERIGMATA PROTEIN.  
GN APSA.  
OS Emericella nidulans (Aspergillus nidulans).  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiiales; Trichocomaceae; Emericella.  
OX NCBI\_TaxID=5072;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN-FGSC 4;  
RX MEDLINE=95164553; PubMed=7860626;  
RA Fischer R., Timberlake W.E.;  
RT "Aspergillus nidulans apsa (anucleate primary sterigmata) encodes a  
RT coiled-coil protein required for nuclear positioning and completion  
RT of asexual development.";  
RN J. Cell Biol. 128:485-498(1995).  
[2]  
RP CHARACTERIZATION  
RX MEDLINE=96025472; PubMed=9379904;  
RA Stelmann R., Slevens N., Fischer R.;  
RT "Nuclear traffic in fungal hyphae: in vivo study of nuclear migration  
RT and positioning in Aspergillus nidulans.";  
CC Mol. Microbiol. 25:757-769(1997).  
CC -1- FUNCTION: REQUIRED FOR NUCLEAR POSITIONING AND COMPLETION OF  
CC ASEXXUAL DEVELOPMENT.  
CC -1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED.  
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X82289; CAA57733.1; -  
DR InterPro; IPR001849; PH.  
DR Pfam; PF00169; PH; 1.  
DR SMART; SM00233; PH; 1.  
DR PROSITE; PS50003; PH\_DOMAIN; 1.  
KW Coiled coil; Membrane  
FT DOMAIN 51 127 COILED COIL (POTENTIAL).  
FT DOMAIN 193 359 COILED COIL (POTENTIAL).  
FT DOMAIN 408 453 COILED COIL (POTENTIAL).  
FT DOMAIN 1393 1504 PH.  
SQ SEQUENCE 1676 AA; 183671 MW; 8BDAL17F4212AEAD8 CRC64;  
  
Query Match 3.9%; Score 123.5; DB 1; Length 1676;  
Best Local Similarity 20.6%; Pred. No. 7.4;  
Matches 139; Conservative 73; Mismatches 222; Indels 241; Gaps 30;  
  
OY 19 GSSEDKMVSAAAGEGEVDELLALGYKVRASDMADVAQRLEOLEMAMGCGVGAAP 78  
DB 936 GTSTIDADVSAISSEQIEPEVEPERK---QDVAIYAEAVPEGPLSF---VEQGTNT 988  
OY 79 DD--SFAT-----HLATDVHYNPDTLSS----- 100  
DB 969 DVEISFPAISSEVEREPAVARETKDVEPEYLSLLEQGTSTDTVEFSVSSISEPEV 1048  
OY 101 -----WVESMSELNAP-----PPP-----LPPAPQLNASTSST 129  
DB 1049 EPIREAKEEAAVDDVASESTHPIFLTPAYTEPTAPKLOEAVIEPPAPQLALSTVSS 1108  
OY 130 V-----TSGGCVFDLPVSDSSSIYALRPIPPAGATAPADASA-----DSVRD-- 174  
DB 1109 VETPPVQYTPDVLILPTPALDENTP-----PSVMASTAKATASAPPLVYVDNTDKG 1161  
OY 175 -----PKRMRTGSGSTSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 203  
DB 1162 TADGLVTQONGVTPLGAIISGNAAPRRARS--GSSNQADQGAQTILSSQIDQLIDRASV 1220  
OY 204 VEAAPVAAANATPALPVYVVDVTOEAGIRLVHALLACAEAVQGNLSAELVYQIPL 263  
DB 1221 RPLSPDSDKINEMNSP-----FATPKARSRVPVPAASNLKRRGS 1263  
OY 264 AASGGAMRKVAAYGEALARRVFRPPOSSLDAFADLLHAFESCPYLFAFHT 323  
DB 1264 AASQASSVQ-----IH-----PPLPADH-- 1281

```

OY 324 ANQALAEAFACGRVHVVDGKIGQMOPALLQALALRPGPPSRLTGVGPPOPETDA 383
DB 1282 -----KEALMAAEKXSIDORPASAGMGPPLPASAVRASSOOR-----ERTPMES-A 1328
OY 384 LQOVGNKLAQFAHTTRVD--FOYRGLVAATLADLEP-FMLQEGEEDPEEPEIAVANSV 440
DB 1329 LQVGSAKTTTSRASYRDSHMSRSSVSFASLEERFNMQP-----NPPFAQGYSTGTD 1384
OY 441 FEMHRLNLPGLAEVYLG-----TVRAVRPRLYVVEQEAHNSTGTELDRTESLHYTS 494
DB 1385 PRMIAATQ-----TWIGELFMKYYTRRAVSGE-----SNTRRRRYFWVHPYTRL--YW 1432
OY 495 TWEDSLGGSSGGGSEVSGAA--PAAAGTD-OWMSEVYLGRQICNVACEGAE 548
DB 1433 SEHDQOSAKSGSGRKSYSIEAVRVADNPYPPGLHCKSLSEVSGRKIRPT-----AT 1487
OY 549 RTERHETLQWNRNL 563
DB 1488 TSORHET---MFNAL 1499

RESULT 12
TRIO_HUMAN STANDARD: PRT: 3038 AA.
AC 075962: Q13458;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE TRIPLE FUNCTIONAL DOMAIN PROTEIN (PTPRF INTERACTING PROTEIN).
GN TRIO.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE-Fibroblast;
RX MEDLINE=96224308; PubMed=8643598;
RA Debant A., Serra-Pages C., Seipel K., O'Brien S., Tang M., Park S.-H.,
RA Streuli M.;
RA "The multidomain protein Trio binds the LAR transmembrane tyrosine
RT phosphatase, contains a protein kinase domain, and has separate rac-
RT specific and rho-specific guanine nucleotide exchange factor
RT domains";
RL Proc. Natl. Acad. Sci. U.S.A. 93:5466-5471(1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Streuli M.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [3]
RP STRUCTURE BY NMR OF 1227-1407, AND MUTAGENESIS.
RX MEDLINE=99005194; PubMed=9790533;
RA Liu X., Wang H., Eberstadt M., Schnuchel A., Olejniczak E.T.,
RA Meadows R.P., Schkeryantz J.M., Janowick D.A., Harlan J.E.,
RA Harris E.A.S., Staunton D.E., Pesik S.W.;
RT "NMR structure and mutagenesis of the N-terminal Dbl homology domain
RT of the nucleotide exchange factor Trio.";
RL Cell 95:269-277(1998).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=99276567; PubMed=10341202;
RA Seipel K., Medley Q.G., Kedersha N.L., Zhang X.A., O'Brien S.P.,
RA Serra-Pages C., Hemler M.E., Streuli M.;
RT "Trio amino-terminal guanine nucleotide exchange factor domain
RT expression promotes actin cytoskeleton reorganization, cell migration
RT and anchorage-independent cell growth.";
RL J. Cell Sci. 112:1825-1834(1999).
RN [5]
RP FUNCTION: PROMOTES THE EXCHANGE OF GDP BY GTP. TOGETHER WITH
RN LENDOCYTE ANTIGEN-RELATED (LAR) PROTEIN, IT COULD PLAY A ROLE IN
RN COORDINATING CELL-MATRIX AND CYTOSKELETAL REARRANGEMENTS NECESSARY
RN FOR CELL MIGRATION AND CELL GROWTH.
CC -! SUBUNIT: INTERACT TO FORM A COMPLEX WITH LEUKOCYTE ANTIGEN RELATED
CC PROTEIN.

```

```

CC -! ALTERNATIVE PRODUCTS: 2 ISOFORMS: 1 (SHOWN HERE) AND 2; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -! TISSUE SPECIFICITY: HIGHLY EXPRESSED IN HEART, SKELETAL MUSCLE,
CC BRAIN, PANCREAS, PLACENTA, LIVER, KIDNEY AND LUNG.
CC -! DOMAIN: THE N-TERMINAL DBL/GEF DOMAIN SPECIFICALLY CATALYZES
CC NUCLEOTIDE EXCHANGE FOR RAC1, LEADING TO THE ACTIVATION OF JUN
CC KINASE AND THE PRODUCTION OF MEMBRANE RUFFLES. THE SECOND DBL/GEF
CC DOMAIN IS AN EXCHANGE FACTOR FOR RHOA AND INDUCES THE FORMATION OF
CC STRESS FIBERS.
CC -! PTM: SERINE PHOSPHORYLATED.
CC -! SIMILARITY: BELONGS TO THE RHO/RAC GEF FAMILY.
CC -! SIMILARITY: CONTAINS 2 DBL-HOMOLOGY DOMAINS (DB).
CC -! SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC -! SIMILARITY: CONTAINS 2 PH DOMAINS.
CC -! SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -! SIMILARITY: CONTAINS 4 SPECTRIN REPEATS.
CC -! SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE SER/THR
CC FAMILY OF PROTEIN KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AF091395; AAC43042.1; -
CC EMBL: U42390; AAC34245.1; -
CC HSSP: P04002; IATF.
CC MIM: 601893; -
CC DR InterPro: IPR001251; CRAL_TRIO.
CC DR InterPro: IPR000719; Euk_pkinase.
CC DR InterPro: IPR003006; Ig_MHC.
CC DR InterPro: IPR003598; Ig_C2.
CC DR InterPro: IPR001849; PH.
CC DR InterPro: IPR000219; RhogEF.
CC DR InterPro: IPR001452; SH3.
CC DR InterPro: IPR002290; Ser_thr_kin_actsite.
CC DR InterPro: IPR002017; Spectrin.
CC DR pfam: PF00047; Ig_1.
CC DR pfam: PF00169; PH; 2.
CC DR pfam: PF00069; pkinase; 1.
CC DR pfam: PF00621; RhogEF; 2.
CC DR pfam: PF00018; SH3; 1.
CC DR pfam: PF00435; spectrin; 7.
CC DR SMART: SM00408; IGC2; 1.
CC DR SMART: SM00233; PH; 2.
CC DR SMART: SM00325; RhogEF; 2.
CC DR SMART: SM00516; SEC14; 1.
CC DR SMART: SM00326; SH3; 2.
CC DR SMART: SM00150; SPEC; 6.
CC DR SMART: SM00220; S_TKC; 1.
CC DR PROSITE: PS50003; PH_DOMAIN_2.
CC DR PROSITE: PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
CC DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
CC DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
CC DR PROSITE: PS50002; SH3; 1.
CC DR Guanine-nucleotide releasing factor: Phosphorylation; Repeat;
CC Transferase; Serine/threonine-protein kinase; ATP-binding;
CC Immunoglobulin domain; SH3 domain; Alternative splicing.
CC REPEAT 252 359
CC REPEAT 479 585
CC REPEAT 819 925
CC REPEAT 1050 1157
CC REPEAT 1240 1393
CC DOMAIN 1421 1532
CC DOMAIN 1597 1653
CC DOMAIN 1917 2091
CC DOMAIN 2098 2212
CC DOMAIN 2630 2707
CC DOMAIN 2737 2993
CC DOMAIN 656 659
CC POLY-GLN.

```

```
FT DOMAIN 1786 1791 POLY-SER.
FT DOMAIN 2233 2253 POLY-GLY.
FT DOMAIN 2486 2492 POLY-SER.
FT DISULFID 2637 2700 POTENTIAL.
FT NP_BIND 2743 2751 ATP (BY SIMILARITY).
FT BINDING 2766 2766 ATP (BY SIMILARITY).
FT ACT_SITE 2856 2856 BY SIMILARITY.
FT VARSPLIC 2242 2308 GSGGGGSGGAPSGSGHSGGAPSGAPSTRSRPSRI
POPVRHHPVIVSSAASQAADKMS -> AAGVCAAAA
GPPVAAATVAAAPAAAAPPARAGAGPGSPSLDTPPCW
SPLOPRAROKTQRO (IN ISOPORM 2).
FT VARSPLIC 2309 3038 MISSING (IN ISOPORM 2).
FT MUTAGEN 1240 1240 E->A: 50% DECREASE IN NUCLEOTIDE EXCHANGE
ACTIVITY.
FT MUTAGEN 1244 1244 T->A: 40% DECREASE IN NUCLEOTIDE EXCHANGE
ACTIVITY.
FT MUTAGEN 1330 1330 N->A: NO CHANGE IN NUCLEOTIDE EXCHANGE
ACTIVITY.
FT MUTAGEN 1367 1367 V->A: 90% DECREASE IN NUCLEOTIDE EXCHANGE
ACTIVITY.
FT MUTAGEN 1368 1368 Q->A: 80% DECREASE IN NUCLEOTIDE EXCHANGE
ACTIVITY.
FT MUTAGEN 1369 1369 R->A: 80% DECREASE IN NUCLEOTIDE EXCHANGE
ACTIVITY.
FT MUTAGEN 1371 1371 T->A: 80% DECREASE IN NUCLEOTIDE EXCHANGE
ACTIVITY.
FT MUTAGEN 1372 1372 K->A: LOSS OF NUCLEOTIDE EXCHANGE
ACTIVITY.
FT MUTAGEN 1375 1375 L->A: 40% DECREASE IN NUCLEOTIDE EXCHANGE
ACTIVITY.
FT MUTAGEN 1378 1378 K->A: NO CHANGE IN NUCLEOTIDE EXCHANGE
ACTIVITY.
FT MUTAGEN 1379 1379 E->A: 30% DECREASE IN NUCLEOTIDE EXCHANGE
ACTIVITY.
SQ SEQUENCE 3038 AA: 341611 MW: 28620F3B513EB74B CRC64;
```

Query Match 3.9%, Score 123.5; DB 1; Length 3038;  
Best Local Similarity 23.7%; Pred. No. 16;  
Matches 58; Conservative 24; Mismatches 74; Indels 89; Gaps 10;

```
QY 4 EYODAGSGGGGGCGGSSSEDKMMVSAAGEFEVDLLAALGYKVRADMDVNAKQLQL 63
DB 2226 EYORNHSGGGGGGSGG-----GGGG----- 2247
QY 64 EMAMGCGVGA-----GAAPDPSFA--THLATDTVHNPTDLSWVESM-- 105
DB 2248 -----GSGGCGAPSGSGSHSGGSCGAPSTRSRPSRIPOPVRHHPVIVSSAASQAE 2303
QY 106 ---LSELNAPPEPLP---APOLNASTSTVTYGGSGYFDLPSPVSSSSITALRIPSPA 159
DB 2304 ADKMSGTSTPGPSLPPGAPAPAGPSAPSR-----RPGDADEGSEEREAEPIPKMK 2354
QY 160 GATAPADLSADSVRPKMKRMKRGSGSTSSSSSSSLG-----CGAASSVVEAPAPVAAAAN 215
DB 2355 VLESRKKAANA-----SGSSPDAPAKDARASLTGLPKPRAGA--ASPLNSPLSS 2404
QY 216 AVPAL 220
DB 2405 AVPSL 2409
RESULT 13
CENB_CRIGR STANDARD: PRT: 606 AA.
AC P46988;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE MAJOR CENTROMERE AUTOANTIGEN B (CENTROMERE PROTEIN B) (CENP-B).
GN CENPB.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
```

```
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetus.
OX NCBI_TaxID=10029;
RN [1]
RC SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=96234058; Pubmed=8652663;
RA Bejarano L.A., Valdivia M.M.;
RT "Molecular cloning of an intronless gene for the hamster centromere
antigen CENP-B."
RL Blochim. Biophys. Acta 1307:21-25(1996).
CC -1- FUNCTION: INTERACTS WITH CENTROMERIC HETEROCHROMATIN IN
CC CHROMOSOMES AND BINDS TO A SPECIFIC SUBSET OF ALPHOID SATELLITE
CC DNA, CALLED THE CENP-B BOX. MAY ORGANIZE ARRAYS OF CENTROMERE
CC SATELLITE DNA INTO A HIGHER ORDER STRUCTURE WHICH THEN DIRECTS
CC CENTROMERE FORMATION AND KINETOCHORE ASSEMBLY IN MAMMALIAN
CC CHROMOSOMES (BY SIMILARITY).
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: STRONG, WITH OTHER MAMMALIAN CENP-B.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL: U20951; AAB06494.1; -.
DR Chromosomal protein; Nuclear protein; DNA-binding; Centromere.
FT DNA_BIND 1 125
FT DOMAIN 404 471 GLU-RICH (ACIDIC).
FT DOMAIN 510 545 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 606 AA: 66407 MW: 63D0EEB551D48E CRC64;
```

Query Match 3.9%, Score 123; DB 1; Length 606;  
Best Local Similarity 19.9%; Pred. No. 2.2;  
Matches 119; Conservative 60; Mismatches 226; Indels 194; Gaps 22;

```
QY 35 EEVDELLAALGYKVRASOM--ADVAKLEOLEEMAMGCGVGAAPDPSFATHLATDTVH 92
DB 75 DKLEGLLLAMFOQIIRAAGLPVKGIILKEKALRIAEELG-----MDPFTAS----- 119
QY 93 YNPITDLSWVESM-----LSELNAPPEPLPAPOLNASTSTVTGSGYF-- 137
DB 120 -----NCMLDRFRRHGVVACSGVTRSRARTSTPRAPAAPAGPAVPS--EGSGGSTPG 171
QY 138 -----DLPSPVDSSTYALRIPSPAGATAPADLSADSVRPKMKRMKRGSGSTSSSSSS 192
DB 172 WTRREOPPVSVEG--YASOVFS--ATETSLWYFLSDQASGLMGSGDGTAKQATQR 224
QY 193 SLSGCGASSVVEAPVAAAANATPALPVVVDTQEGAIRLVHALLACAEVQOENLSA 252
DB 225 LSVLLCANRDGSEKPLPVAGKSAPR-----RASGGGLCDITYANKSGVY 270
QY 253 AEALVKQIPLLAASOGAMRKVAAYFGEALARVFRFRPDPSSLLDAFADLHAHYE 312
DB 271 TQALAKYLKALDTRMAASRRVLLAGRLAOSL-----DYS-----GLRHVQL-- 314
QY 313 SCPYLKFAHFTANQAILLEFACGRVHYVDGFIKGM-----QMPALLQALALRGPSPSF 368
DB 315 -----AFPPPGTVHPLERGVVQOVGVHQAQMLKAMALEGQDPSG 356
QY 369 RUTGVPPQPDDETALQY--GKTLAQFAHTIRVDFOYRGVLAATLADLEPMLQPEGE 426
DB 357 LQLG-----VEALHVVAAAMQAVEPR-DIATCFREAGFGCGGNATITTT-SFSSEGE 407
QY 427 DNEPEPVIAVSVEMHRLLAQPALEKVLGTVAVPRIYTVVEQEAHNHSGTFDLRF 486
DB 408 EEEEEE-----EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE 429
```

QY 487 TESLHYSTMFDLSGGSGGSPSEVSSGAAAPAAAGTDQVMSEVYLGRQICNVACG 546  
 DB 430 EE-----EEGGEEGEGEEV-----GEEVEE-----EG 455  
 OY 547 AERTRHETLGGRRNLGNAGFEYVHLGSNAVKQASTLALFAGDGYVEKEGCLT 605  
 DB 456 DESDEEEEEESESSSEGLE-----AEDWAGVVEASGFGGYSVGEACPTL 507

RESULT 14  
 A180\_MOUSE STANDARD; PRT; 901 AA.  
 ID A180\_MOUSE 061547; 061547; 061547; 061547; 061547; 061547; 061547; 061547; 061547; 061547;  
 AC 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE CLATHRIN COAT ASSEMBLY PROTEIN APL180 (CLATHRIN COAT ASSOCIATED PROTEIN APL180) (PHOSPHOPROTEIN F1-20) (91 KDA SYNAPTOSOMAL-ASSOCIATED PROTEIN).  
 DE SNAP91.  
 GN SNAP91.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP MEDLINE=92300439; PubMed=1607933;  
 RA Zhou S., Sousa R., Tannery N.H., Lafer E.M.;  
 RT "Characterization of a novel synapse-specific protein. II. cDNA cloning and sequence analysis of the F1-20 protein.";  
 RL J. Neurosci. 12:2144-2155(1992).  
 CC -1- FUNCTION: ADAPTINS ARE COMPONENTS OF THE ADAPTOR COMPLEXES WHICH LINK CLATHRIN TO RECEPTORS IN COATED VESICLES. CLATHRIN-ASSOCIATED PROTEIN COMPLEXES ARE BELIEVED TO INTERACT WITH THE CYTOPLASMIC TAILS OF MEMBRANE PROTEINS, LEADING TO THEIR SELECTION AND CONCENTRATION. BINDING OF APL180 TO CLATHRIN TRISKELIA INDUCES THEIR ASSEMBLY INTO 60-70 NM COATS.  
 CC -1- SUBCELLULAR LOCATION: COMPONENT OF THE COAT SURROUNDING THE CYTOPLASMIC FACE OF COATED VESICLES IN THE PLASMA MEMBRANE.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: BRAIN. ASSOCIATED WITH THE SYNAPSES.  
 CC -1- DEVELOPMENTAL STAGE: DEVELOPMENTALLY REGULATED IN A PATTERN COINCIDENT WITH ACTIVE SYNAPTOGENESIS AND SYNAPTIC MATURATION.  
 CC -1- DOMAIN: POSSESSES A THREE DOMAIN STRUCTURE: THE N-TERMINAL 300 RESIDUES HARBOUR A CLATHRIN BINDING SITE, AN ACIDIC MIDDLE DOMAIN 450 RESIDUES, INTERRUPTED BY AN ALA-RICH SEGMENT, AND THE C-TERMINAL DOMAIN (166 RESIDUES).  
 CC -1- PTM: PHOSPHORYLATED.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; M83985; AAA37587.1; -;  
 DR EMBL; M83985; AAA37586.1; -;  
 DR HSSP; P04002; 1MFA.  
 DR MGD; MGI:109132; Snap91.  
 DR InterPro: IPR001026; ENTH.  
 DR Pfam; PF01417; ENTH; 1.  
 DR SMART; SM00273; ENTH; 1.  
 KW Coated pits; Alternative splicing; phosphorylation.  
 FT DOMAIN 410 413 POLY-THR.  
 FT DOMAIN 535 539 POLY-ALA.  
 FT DOMAIN 547 550 POLY-ALA.  
 FT DOMAIN 659 664 POLY-SER.  
 FT DOMAIN 704 710 POLY-SER.  
 FT VARSPIC 715 719 MISSING (IN SHORT ISOFORM).  
 SQ SEQUENCE 901 AA; 91851 MW; 24A98FBACE8DB8B1 CRC64;

Query Match 3.8%; Score 122.5; DB 1; Length 901;  
 Best Local Similarity 22.7%; Pred. No. 3.9; Mismatches 212; Indels 181; Gaps 28;  
 Matches 132; Conservative 57; Mismatches 212; Indels 181; Gaps 28;

OY 50 ASDMDVAOKLEOLEMAOMGGVAGAA-----PDSFATH-----LATDVHYNPTDLSM 101  
 DB 337 AAPVSAKPSDLIDQPPFSGAAGAAAPVPPSGATRAMGLDGLDGL-----AALSSV 392  
 OY 102 -VESMLSEINAPPPPLP-----PAPQLNASTSTVTGSGGYFDLPVSDDSSSYVALRPI 155  
 DB 393 PCFAPISDPEFADPEPSPPTTTTBP-SASASTTAYVA-----VTTEVDLFGDFAFASPG 445  
 OY 156 PSRA-----GATAPAD-----LSADSVRPKRRKRGSGSSSSSSSSSSSLGGARSSVE 205  
 DB 446 EAPAASEGATAPTPAPVAALDACSGNDP-----FASESGAEAPELDLFAMPPEP 499  
 OY 206 AAPVAAANATPAPLVVVVDFQEAGIRLVHA-LLACAAVOGENTSAEALVKQIP--- 261  
 DB 500 SAPVYPTASTAPPVATAPSPAPTAATAATTTTAAATTTATTSAAATTAAPRAL 559  
 OY 262 -----LLAASGGAMKRYAAYFGELARRVFRPQDSSLDAFADLLHAH 309  
 DB 560 DIFGDLFDSAPVEVAAPKPDAPSIDLFGTDAFSSPPRGASPVPESSLT---ADLLSYD 615  
 OY 310 FYESCPLYLFAHTANQALLLEAFAGCRVHYVD-FGICQMOPALLQALALRPGGPSE 368  
 DB 616 AFPA-----PSPASTASPAAESG-----VIDLFGDARG----- 645  
 OY 369 RLTGVPQPDDELDLQOYGMWKAQFAHTIRVDFOYRGVLAATLADL-----EPFLQPE 423  
 DB 646 --SGASETP--APQAVSSSAS-----ADLLAGFGGSFMA--- 677  
 OY 424 GEEDPNEPEVIAVNSVFEMHRLAOPGALKEVLTGVAVRPRITVVBQEAHNHNGFTL 483  
 DB 678 ----PSTTVTPAQN-----LQRP-SPEAFAFGTTPSR----- 705  
 OY 484 DRFTESLHYSTMFDLS-----EGSSGGGSPSEVSS-----SGAAAPAAAGTDQVMSEVYL 534  
 DB 706 ---SSSSSPDSVFDELGLDLPMTNAPSGQPAVSWPPSPMAAASKGLSD-----L-755  
 OY 535 GRQICNVACEGAEERTHERHTLGGRRNLGNAGFEYVHLGSN 576  
 DB 756 DSSLASLVGNLIGSTTSKKGLDQ-----NAGEKRLTGGAN 792

RESULT 15  
 CYAA\_USTMA STANDARD; PRT; 2493 AA.  
 ID CYAA\_USTMA 01-FEB-1996 (Rel. 33, Created)  
 AC P49606;  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE ADENYLATE CYCLASE (EC 4.6.1.1) (ATP PYROPHOSPHATE-LYASE) (ADENYLATE CYCLASE).  
 DE UACT OR REM1.  
 GN Ustilago maydis (Smut fungus).  
 OS Ustilago maydis (Smut fungus).  
 OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes; Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.  
 OX NCBI\_TaxID=5270;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=518;  
 RX MEDLINE=95087882; PubMed=7995519;  
 RA Gold S., Duncan G., Barrett K., Kronstad J.W.;  
 RT "CAMP regulates morphogenesis in the fungal pathogen Ustilago maydis.";  
 RL Genes Dev. 8:2805-2816(1994).  
 CC -1- FUNCTION: PLAYS ESSENTIAL ROLES IN REGULATION OF CELLULAR METABOLISM BY CATALYSING THE SYNTHESIS OF A SECOND MESSENGER, CAMP.  
 CC -1- CATALYTIC ACTIVITY: ATP -> 3',5'-CYCLIC AMP + PYROPHOSPHATE.  
 CC

CC -1- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-3 FAMILY.  
CC -1- SIMILARITY: CONTAINS 21 LEUCINE-RICH REPEATS (LRR).  
CC -1- SIMILARITY: CONTAINS A PP2C-LIKE DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: L33918; AA57469.1; -  
DR InterPro: IPR001054; Guanylt\_cyclase.  
DR InterPro: IPR001611; LRR.  
DR InterPro: IPR003592; LRR\_out.  
DR InterPro: IPR003591; LRR\_tyr.  
DR InterPro: IPR003589; PP2C\_catalytic.  
DR InterPro: IPR001932; PP2C\_domain.  
DR InterPro: IPR000159; RA.  
DR Pfam: PF00211; guanylate\_cyc; 1.  
DR Pfam: PF00560; LRR; 15.  
DR Pfam: PF00481; PP2C; 1.  
DR SMART: SM00044; CYC; 1.  
DR SMART: SM00370; LRR; 6.  
DR SMART: SM00369; LRR\_tyr; 2.  
DR SMART: SM00332; PP2C; 1.  
DR SMART: SM00314; RA; 1.  
DR PROSITE: PS50125; GUANYLATE\_CYCLASES\_2; 1.  
KW Lyase; Repeat; Leucine-rich repeat; CAMP synthesis; Magnesium.  
FT DOMAIN 759 763 POLY-ASP.  
FT REPEAT 888 897 POLY-ALA.  
FT REPEAT 1084 1106 LRR 1.  
FT REPEAT 1108 1131 LRR 2.  
FT REPEAT 1132 1155 LRR 3.  
FT REPEAT 1157 1178 LRR 4.  
FT REPEAT 1179 1201 LRR 5.  
FT REPEAT 1202 1225 LRR 6.  
FT REPEAT 1227 1248 LRR 7.  
FT REPEAT 1249 1271 LRR 8.  
FT REPEAT 1273 1295 LRR 9.  
FT REPEAT 1314 1337 LRR 10.  
FT REPEAT 1338 1360 LRR 11.  
FT REPEAT 1361 1384 LRR 12.  
FT REPEAT 1386 1407 LRR 13.  
FT REPEAT 1408 1430 LRR 14.  
FT REPEAT 1432 1455 LRR 15.  
FT REPEAT 1509 1529 LRR 16.  
FT REPEAT 1533 1556 LRR 17.  
FT REPEAT 1557 1580 LRR 18.  
FT REPEAT 1581 1604 LRR 19.  
FT REPEAT 1606 1628 LRR 20.  
FT REPEAT 1633 1656 LRR 21.  
FT DOMAIN 1722 2001 PP2C-LIKE.  
FT DOMAIN 2002 2493 CATALYTIC.  
SQ SEQUENCE 2493 AA; 271979 MW; 106A872C3C1C5BF8 CRC64;

Query Match 3.7%; Score 119.5; DB 1; Length 2493;  
Best Local Similarity 23.2%; Pred. NO. 21;  
Matches 66; Conservative 41; Mismatches 106; Indels 71; Gaps 12;

QY 6 ODAGSGGGGGG-----MGSEEDKM-----WYSAAGGEEVDLLAL----- 44  
DB 111 QDRSSGSGSGSGSRYPANSHSDSLPSPSTHSISPSFDE-DELROQIMSHIPANQATSS 169  
QY 45 -----GYKVRASDMADVAQLEOLEMAMGCGVGAGAAPDSEFATHLATDVHYNPDTLS 99  
DB 170 SDGDYGRKAVQSANHQDISPFLQSE-----NAAPFSS-----SHSNRTSVN 210  
QY 100 SWESMSELNAPPPPLPPAPOLNASTSYVTGSGGYFDLPSPVDSSSYALARPISPA 159  
DB 211 PSAASTASPSSTAATRTPRGCTNASTQYNTLDTSPGSDID-RPGLSSRSRQYSLRPQTPPS 269

QY 160 GATAPADLSAD-----SVRDP-----KRMRTGSGSTSSSSSSSISGGGARSSVVEA 206  
DB 270 ASTSTSTLNGSKDTHASAVKKTRNPFGLKKSSAHSNASSNHPTRHDI--GSVSSLSR 327  
QY 207 APPVAAAANATPALFVVYVDIOEACIRLYVHALLACAEVQOENL 250  
DB 328 YGP-NAANVNPMPRPAPWLDN-----HCTLANSNSPSSASTL 362

Search completed: January 22, 2002, 16:30:28  
Job time: 338 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

```
Run on:      January 22, 2002, 16:24:25 ; Search time 40.9 seconds
              (without alignments)
              2228.060 Million cell updates/sec
```

Title: US-09-485-529-7  
Perfect score: 3194  
Sequence: 1 MKREYQDAGSGSGGGGGMGS.....TLGNHTPLIATSAWRLAP 6233

Scoring table:      BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues  
Total number of hits satisfying chosen parameters: 473505

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```

1:  sp.archaea:*
2:  sp.bacteria:*
3:  sp.fungi:*
4:  sp.human:*
5:  sp.invertebrate:*
6:  sp.mammal:*
7:  sp.mhc:*
8:  sp.organelle:*
9:  sp.phage:*
10: sp.plant:*
11: sp.rodent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3194	100.0	623	10	Q9ST59	Q9ST59 triticum aea
2	2683	84.0	630	10	Q9ST48	Q9ST48 zea mays (m
3	2669	83.6	625	10	Q9MB96	Q9MB96 oryza sativa
4	1612	50.5	587	10	Q23642	Q23642 arabidopsis
5	1608	50.3	587	10	Q9SLH3	Q9SLH3 arabidopsis
6	1604	50.2	587	10	Q23725	Q23725 arabidopsis
7	1594	49.9	532	10	Q23724	Q23724 arabidopsis
8	1594	49.9	533	10	Q9LQ78	Q9LQ78 arabidopsis
9	1590	49.8	532	10	Q23643	Q23643 arabidopsis
10	1530.5	47.9	547	10	Q9SRP9	Q9SRP9 arabidopsis
11	1511.5	47.3	511	10	Q9C8V3	Q9C8V3 arabidopsis
12	1495.5	46.8	662	10	Q65367	Q65367 arabidopsis
13	1450.5	45.4	523	10	Q9LFE3	Q9LFE3 arabidopsis
14	1005.5	31.5	493	10	Q9ASV7	Q9ASV7 oryza sativa
15	670	21.0	668	10	Q9FUF7	Q9FUF7 zea mays (m
16	606	19.0	617	10	Q9LWU9	Q9LWU9 oryza sativa
17	604.5	18.9	653	10	Q9M384	Q9M384 arabidopsis
18	600.5	18.8	653	10	Q96304	Q96304 arabidopsis
19	599	18.8	428	10	Q9ZSP2	Q9ZSP2 lycopersicon

20	587.5	18.4	490	10	Q9L1L7	Q9L1L7 arabidopsis
21	585.5	18.3	819	10	Q9AVK4	Q9AVK4 pisum sativ
22	572.5	17.9	482	10	Q9LPR8	Q9LPR8 arabidopsis
23	558	17.5	526	10	Q9LPT0	Q9LPT0 arabidopsis
24	554	17.3	413	10	Q9S7H5	Q9S7H5 arabidopsis
25	551	17.3	658	10	Q9CANA3	Q9CANA3 arabidopsis
26	535	16.8	445	10	Q9ZWC5	Q9ZWC5 arabidopsis
27	527.5	16.5	584	10	Q9PFL03	Q9PFL03 arabidopsis
28	526	16.5	447	10	Q9ARF6	Q9ARF6 capsella ru
29	523.5	16.4	405	10	Q9FHZ1	Q9FHZ1 arabidopsis
30	512	16.0	313	10	Q9ZTB1	Q9ZTB1 zea mays (m
31	507	15.9	306	10	Q9XE53	Q9XE53 arabidopsis
32	498	15.6	593	10	Q9SDQ3	Q9SDQ3 arabidopsis
33	464	14.5	352	10	Q9XE51	Q9XE51 arabidopsis
34	462	14.5	1502	10	Q9LXNK6	Q9LXNK6 arabidopsis
35	445.5	13.9	1336	10	Q9L1W74	Q9L1W74 arabidopsis
36	444.5	13.9	542	10	Q9SCRO	Q9SCRO arabidopsis
37	443	13.9	808	10	Q9XE58	Q9XE58 arabidopsis
38	439	13.7	610	10	Q9LIT5	Q9LIT5 arabidopsis
39	433.5	13.6	551	10	Q9SEF7	Q9SEF7 arabidopsis
40	429.5	13.4	284	10	Q9XSE7	Q9XSE7 arabidopsis
41	415	13.0	375	10	Q93566	Q93566 arabidopsis
42	402.5	12.6	325	10	Q9XE52	Q9XE52 arabidopsis
43	399.5	12.5	718	10	Q80933	Q80933 arabidopsis
44	399	12.5	287	10	Q9NM05	Q9NM05 arabidopsis
45	396	12.4	483	10	Q9S0F5	Q9S0F5 arabidopsis

## ALIGNMENTS

RESULT	1
ID	09ST59
AC	09ST59 PRELIMINARY; PRT: 623 AA.
DT	01-MAY-2000 (TREMBLrel. 13, Created)
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT	01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE	GIBBERELLIN RESPONSE MODULATOR.
GN	RHT-DIA.
OC	Triticum aestivum (Wheat).
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
CC	Triticeae; Triticum.
OX	NCBI_TaxID=4565;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=99347734; PubMed=10421366;
RA	Peng J., Richard D.E., Hartley N.M., Murphy G.P., Devos K.M.,
RA	Flintham J.E., Beales J., Fish L.J., Worland A.J., Pelica F.,
RA	Sudhakar D., Christou P., Snape J.W., Gale M.D., Harberd N.P.;
RT	"Green Revolution' genes encode mutant gibberellin response
RT	modulators.";
RL	Nature 400:256-261(1999).
DR	EMBL: AJ242531; CAB51555.1; -
SQ	SEQUENCE 623 AA; 65337 MW; F2BAC34996D0A84F CRC64;

Query Match	100.0%; Score 3194; DB 10; Length 623;
Best Local Similarity	100.0%; Pred. No. 8.3e-199;
Matches 623; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	

Dy	1 MKREVDAGGSGGGGGSSFDKMMVSAAGESEVEVELLAIGYKVVASMDADVAOKL 60
Dz	
Dc	1 MKREFODAGSGSGGGGSSDDKMMVSAAGESEVEVELLAIGYKVVASMDADVAOKL 60
Dd	
Dy	61 EOLENAMMGVGGAAPADDSEFATHLATDVTHYNPTDLSSWVESMLSELNAPPPLPPAP 120
Dz	
Dc	61 EOLENAMMGVGGAAPADDSEFATHLATDVTHYNPTDLSSWVESMLSELNAPPPLPPAP 120
Dd	
Dy	121 QLNASTSYTVGSGGYFDLPSPVSDDSSSIYALRPIPSPAGATAPADLSADVDRDKRMET 180
Dz	
Dc	121 QLNASTSYTVGSGGYFDLPSPVSDDSSSIYALRPIPSPAGATAPADLSADVDRDKRMET 180
Dd	

[illegible][illegible]

Qy	117	P--PAPDNNSTSYT---GSGGFEDLPPEYDSSSTYALRPIPSGAPADLSL	169
Db	118	PPAPPAPL-ASTSYTSGAAGAGTFDLPRAVDSSSTYALKRIPSPMAAS-ADPS	175
Qy	170	DSYDPKRMRTGSGSTSSSSSSSSSSSGG--ARSSVVEAEP--VAANAATPALPVVY	225
Db	176	DSAREPKRMRTGSGSTSSSSSSSSSMGGRTRSSVVEAAPATQASMAANG--PAVPVVY	234
Qy	226	DTQENGIRLVALLACAEVAQOENLSAAELVKOITPLAASOGAMKVAAYGELAR	285
Db	235	DTQENGIRLVALLACAEVAQOENFSAEALVKOITPLAASSOGAMKVAAYGELAR	294
Qy	286	VFRFRPODSSLDAAPADLLHAHFYESCPLKFAHFTANQAILLEAFAGCRHVHVDGI	345
Db	295	VYRFRPPDSSLDAAPADLLHAHFYESCPLKFAHFTANQAILLEAFAGCRHVHVDGI	354
Qy	346	KQGMOWPALLOALALRGGPSPFLTGVPQRPETALLOQVGKLAQAFHTLRVDFQYR	405
Db	355	KQGMOWPALLOALALRGGPSPFLTGVPQRPETALLOQVGKLAQAFHTLRVDFQYR	414
Qy	406	GLVATTLADLEPFMLQEGEDRPEREEVIANVSVFEMHLLAQPGALEKVLGTVAVRP	465
Db	415	GLVATTLADLEPFMLQEPG--DQTDDEPVIIVANVSVELHRLLAQPGALEKVLGTVAVRP	473
Qy	466	RIVYVVEQEAHNNNGSTFLDRTEBSLHYSTMPDLSBEGSSGGGSEVSSGAAAPAAAGT	525
Db	474	RIVYVVEQEAHNNNGSTFLDRTEBSLHYSTMPDLSBEGAGSGSTDAAP---AAGGT	529
Qy	526	DQVMSVYLGRQIKNVVAECGAERTERHETLIGQMRNIGNAGSEPTVHLLGSNAYKOASTLL	585
Db	530	DQVMSVYLGRQIKNVVAECGAERTERHETLIGQMRSLGSGGFAPVHLLGSNAYKOASTLL	589
Qy	586	ALFAGGDGYKVEEKBGCCTLJGWHRRPLIATISAMKLA	621
Db	590	ALFAGGDGYRVEEKDGCCTLJGWHRRPLIATISAMKVA	625

RESULT	3			
Q9MB96		PRELIMINARY;	PRT;	625 AA.
AC	Q9MB96.			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)			
DE	OSGAI.			
GN	OSGAI.			
OS	Oryza sativa (Rice).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;			
OC	Ehretidoideae; Oryzeae; Oryza.			
OX	NCBI_TaxID=4530;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV, NIPONBARE;			
RX	MEDLINE=20179680; PubMed=10713441;			
RA	Ogawa M., Kusano T., Katsumi M., Sano H.;			
RT	"Rice gibberellin-insensitive gene homolog, OSGAI, encodes a nuclear-			
RT	localized protein capable of gene activation at transcriptional			
RT	level.";			
RL	Gene 245:21-29(2000).			
DR	EMBL; AB030956; BAA90749.1; -			
SO	SEQUENCE 625 AA; 65406 MW; 034FF02719D42E97 CRC64;			

[illegible]

QY	59	KLEOLEAMAGGAGGAGAGAPDPSATPLADDTYVNTDLSNWSWESLSELMAPPLP	110
Db	60	KLEOLEAMAGGAGVSAQPADDCGVSHLADDTYVNTDLSNWSWESLSELMAPPLP	119
QY	119	APOL--NASTSVTGTSG--GYFDLPSPVSSSVSLVALRPIPSGATAPADISADV	174
Db	120	APPARHASTSVTGTGGGGSGFFELPAAADSSSVTLARPIISLPVATADPS--AADSARD	178
QY	175	PKRRRTGGSSSTSSSSSSSSSLGGGA--RSSVYEAPV---AAAATPALPVVVTDTQEA	230
Db	179	TKRRRTGGSGTSSSSSSSSSLGGGASGVSVEAAPPTOGAAAAANA--PAVPVVVDTQEA	237
QY	231	GIRLVHALLACAEVVOOENISAAALKQJPLLAASOGGAMRKVAAVGEALARRVFR	290
Db	238	GIRLVHALLACAEVVOOENRFAAALVKQJPLLAASOGGAMRKVAAVFGELARRVFR	297
QY	291	PODPSLLDAAFADLLAHAFYESCPYLKFAHETANOALAEAFAGCRRHVHVDGIRKQMG	350
Db	298	P-ADSTLLDAAFADLLAHAFYESCPYLKFAHETANOALILEAFAGCHRHVHVDGIRKQMG	356
QY	351	WPALLQALALRPGSPSPFRLLTVGVPPOPDDELALQOYGKMLAQFAHTIIRDFQIRGLVAA	410
Db	357	WPALLQALALRPGSPSPFRLLTVGVPPOPDDELALQOYGKMLAQFAHTIIRDFQIRGLVAA	416
QY	411	TLADLEPMLOPBEGEEDNPPEEPIAVANSYFEMBRLLAOGALEKVTGTVRAVPRIVTV	470
Db	417	TLADLEPMLOPBEEADANEPEPIAVANSYFELHRLLAOGALEKVTGTVHAAVPRIVTV	476
QY	471	VEOBANHNSTGFLDRFTESLHYVSTYMDSELEGSSGGGSPSEVSSGAAAAPAAAGTDQVMS	530
Db	477	VEOBANHNSTGFLDRFTESLHYVSTYMDSELEGSS--GQALESPAPAG--GGGSTDQVMS	532
QY	531	EVLIGRQICNVVACEGAEARTRRHETLQWRNRRLGNAGFEYVHIGSNAYKQASTLLAFAG	590
Db	533	EVLIGRQICNVVACEGAEARTRRHETLQWRNRRLGAGFEYVHIGSNAYKQASTLLAFAG	592
QY	591	GDGKVEKEECCTLGMHTRPLINTSAMRLA 621	
Db	593	GDGKVEKEECCTLGMHTRPLINTSAMRYA 623	
RESULT	4		
ID	023642	PRELIMINARY;	PRT; 567 AA.
AC	023642;		
DT	01-JAN-1998 (Tremblrel. 05, Created)		
DT	01-JAN-1998 (Tremblrel. 05, last sequence update)		
DT	01-NOV-1999 (Tremblrel. 12, last annotation update)		
GN	RGAL PROTEIN.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	eucosids II; Brassicales; Brassicaceae; Arabidopsits.		
OX	NCBI_Taxid=3702;		
RP	SEQUENCE FROM N.A.		
RP	MEDLINE=97379310; Pubmed=9237632;		
RA	Truong H.N., Caboche M., Daniel-Vedele F.;		
RT	"Sequence and characterization of two Arabidopsis thaliana cDNAs		
RT	isolated by functional complementation of a yeast gln3 gdh1 mutant."		
RL	FEES Lett. 410:213-218(1997).		
DR	EMBL; Y11336; CAAT2177.1; -		
DR	Mendel; 24145; Arachn; 3051; 24145.		
SO	SEQUENCE 587 AA; 64023 MW; D0A7A3C741PB5IEF CRC64;		

```

Query Match          50.5%;  Score 1612;  DB 10;  Length 587;
Best Local Similarity 53.7%;  Pred. No. 1.8e-96;
Matches 340;  Conservative 82;  Mismatches 149;  Indels 62;  Gaps 13;

0y      1 MKRQYQDAG----SGGGGGGCMGSSSEDKMNV--SAAAGCEVDELALALGKVRASDMA 54
||||:  |      |      |      |      |      |      |      |      |      |

```

[illegible]

99SLH3		RESULT	5
ID	Q9SLH3	PRELIMINARY:	PRT; 587 AA.
AC	Q9SLH3;		
DT	01-MAY-2000 (TREMBLrel. 13, Created)		
DR	01-MAY-2000 (TREMBLrel. 13, Last sequence update)		
DJ	01-MAY-2000 (TREMBLrel. 13, Last annotation update)		
DE	PURATIVE RGAL, GIBBERELLIN RESPONSE MODULATION PROTEIN. AT2G01570.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	eucosids II; Brassicales; Brassicaceae; Arabidopsits.		
OX	Ncrl_TaxId=3702;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-CV. COLOMBIA;		
RX	MEDLINE-20083487; PubMed-10617197;		
RA	Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D., Fuji C.Y., Mason T.M., Bowman J.L., Barnstead M.E., Feldguyum T.V., Buell L.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S. Cronin L.A., Shen M., Vankken S.E., Umayam L., Tallon L.J., Gill J.E. Adams M.D., Carreira A.J., Creasy T.H., Goodman H.M., Somerville C.R., RA Copenhagen G.P., Pleusns D., Nierman W.C., White O., Eisen J.A."		
RA	Salzberg S.L., Fraser C.M., Venter J.C.,"		
RL	"Sequence and analysis of chromosome II of Arabidopsis thaliana.";		
DR	Nature 402:761-768(1999).		
EMBL:	AC005560; AAC67333.1;-		
SQ	SEQUENCE 587 AA; 64035 MW; FC92E7F9408072AA CRC64;		

Query Match 50.3%; Score 1608; DB 10; Length 587;  
 Best Local Similarity 53.6%; Pred. No. 3,3e-96;  
 Matches 339; Conservative 82; Mismatches 150; Indels 62; Gaps 13;

```

OY 1 MKREYQDAG---SGGGGGGMSSEDKMY--SAAAGEBEVDELLAALGYKVRASDMA 54
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1 MKRDHOFGRSLNNGTSSSSSSSSSKDKMMYKKKEEDGGNMDELLAVLGYKVRSSBMA 60
OY 55 DVAOKLEOLEMAMGCGAGAGAPDDSFATHTLATDTVYHNPDLSSWESMSELNAPPP 114
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 61 EVALKLEOLETMM-----SNVOEDGLSHLADTVYHNPSELTSMIDNMLSELN--PP 110
OY 115 PLPPAPQLNASTSVTSGGYFDLPSPVD--SSSIYALRPIPSPAQAPAPA-DLSAD 170
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 111 PLP-----ASNGLDPLVLPPEICGFPASDYDLKVIQGNMIIYQEPALDSSSS 157
OY 171 SVRDPKRRRTGGSSTS--SSSSSSSLGCGARSSVVEAPPYAAANATPALPVVVDQ 228
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 158 SNNOKRRLKSSGSPDSMTSTGTGIGVIGTIVTTTTTTTAAESTRS--VILVDSQ 215
OY 229 EAGIRLVHALLACAAVOEENLSAEALVKOIPLLAASOGAMRKVAAYFGFALARVYR 288
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 216 ENGVALVHALMACARAIQNNITLAEALVKIGCLAVSOGAMRKVAATYFPAALARRIYR 275
OY 289 FRPOPSSLLDAAFADLLHAHFEPCPYLKFAHFTANOALIEAFAGCRRVHVVDGIRG 348
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 276 L--SPQNOIDHCLSDTLQMHFEYETCPYLFKAHFTANOALIEAFGKKRVHVIDESMNG 333
OY 349 MOWPALLOALALRGGPSPFRITGVPQPDETDALOQVGMKLAOPARTIRVDYRGVLY 408
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 334 LQWPAIMQALALREGPPTFRITGIGPAPDSDHLHEVGKLAOLAAVHVEFEYRGFV 393
OY 409 AATLADLEPPMLQPGEDDPNEPEVIAVNSYFEMHRLLAOPGALEKVLGTVAARPRIV 468
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 394 ANSLADLDASML-----ELRPS-DTEAVAVNSYFELHKLGRGGIEKVLGYVKQIKPIYF 448
OY 469 TVVEQEAHNHNSGTFLDRTESLHYSTMFDSLEGSSGGGSEVSSGAAAPAAAGTDQV 528
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 449 TVVEQESHNHNGPVFLDRRTESLHYSTLFDSLEG-----VPSQKQV 490
OY 529 MSEVYLGQIQCNVACBEAERTERHETLQGNRNLGNAGFETVHLGSNAKYQASTLLALF 588
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 491 MSEVYLGQIQCNVACBEADRYERHETLSQWGNRSGSLAPAHILGSNAFKQASMLLSVF 550
OY 589 AGDGYKVEEKEGCLTGLMHTRPPLATASAMRLA 621
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 551 NSGCGYRVEESNGCLMGLMHTRPPLITTSAMKLS 583

```

## RESULT 6

```

ID 023725 PRELIMINARY; PRT: 587 AA.
AC 023725;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE GRS PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Peng J., Carol P., Richards D.E., King K.E., Cowling R.J.,
RA Murphy G.P., Harberd N.P.;
RL Genes Dev. 0:0-0(0).
DR EMBL: Y15194; CAA75493.1;
DR Mendel. 24071; Atach.3051.24071.
SQ SEQUENCE 587 AA; 64006 MW; F6F6C7738EE7DCA9 CRC64;

```

Query Match 50.2%; Score 1604; DB 10; Length 587;  
 Best Local Similarity 53.4%; Pred. No. 6,1e-96;  
 Matches 338; Conservative 82; Mismatches 151; Indels 62; Gaps 13;

```

OY 1 MKREYQDAG---SGGGGGGMSSEDKMY--SAAAGEBEVDELLAALGYKVRASDMA 54
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1 MKRDHOFGRSLNNGTSSSSSSSSSKDKMMYKKKEEDGGNMDELLAVLGYKVRSSBMA 60
OY 55 DVAOKLEOLEMAMGCGAGAGAPDDSFATHTLATDTVYHNPDLSSWESMSELNAPPP 114
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 61 EVALKLEOLETMM-----SNVOEDGLSHLADTVYHNPSELTSMIDNMLSELN--PP 110
OY 115 PLPPAPQLNASTSVTSGGYFDLPSPVD--SSSIYALRPIPSPAQAPAPA-DLSAD 170
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 111 PLP-----ASNGLDPLVLPPEICGFPASDYDLKVIQGNMIIYQEPALDSSSS 157
OY 171 SVRDPKRRRTGGSSTS--SSSSSSSLGCGARSSVVEAPPYAAANATPALPVVVDQ 228
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 158 SNNOKRRLKSSGSPDSMTSTGTGIGVIGTIVTTTTTTTAAESTRS--VILVDSQ 215
OY 229 EAGIRLVHALLACAAVOEENLSAEALVKOIPLLAASOGAMRKVAAYFGFALARVYR 288
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 216 ENGVALVHALMACARAIQNNITLAEALVKIGCLAVSOGAMRKVAATYFPAALARRIYR 275
OY 289 FRPOPSSLLDAAFADLLHAHFEPCPYLKFAHFTANOALIEAFAGCRRVHVVDGIRG 348
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 276 L--SPQNOIDHCLSDTLQMHFEYETCPYLFKAHFTANOALIEAFGKKRVHVIDESMNG 333
OY 349 MOWPALLOALALRGGPSPFRITGVPQPDETDALOQVGMKLAOPARTIRVDYRGVLY 408
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 334 LQWPAIMQALALREGPPTFRITGIGPAPDSDHLHEVGKLAOLAAVHVEFEYRGFV 393
OY 409 AATLADLEPPMLQPGEDDPNEPEVIAVNSYFEMHRLLAOPGALEKVLGTVAARPRIV 468
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 394 ANSLADLDASML-----ELRPS-DTEAVAVNSYFELHKLGRGGIEKVLGYVKQIKPIYF 448
OY 469 TVVEQEAHNHNSGTFLDRTESLHYSTMFDSLEGSSGGGSEVSSGAAAPAAAGTDQV 528
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 449 TVVEQESHNHNGPVFLDRRTESLHYSTLFDSLEG-----VPSQKQV 490
OY 529 MSEVYLGQIQCNVACBEAERTERHETLQGNRNLGNAGFETVHLGSNAKYQASTLLALF 588
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 491 MSEVYLGQIQCNVACBEADRYERHETLSQWGNRSGSLAPAHILGSNAFKQASMLLSVF 550
OY 589 AGDGYKVEEKEGCLTGLMHTRPPLATASAMRLA 621
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 551 NSGCGYRVEESNGCLMGLMHTRPPLITTSAMKLS 583

```

## RESULT 7

```

ID 023724 PRELIMINARY; PRT: 532 AA.
AC 023724;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE GAI PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LANDSBERG ERECTA;
RX MEDLINE=98051192; PubMed=9389651;
RA Peng J., Carol P., Richards D.E., King K.E., Cowling R.J.,
RA Murphy G.P., Harberd N.P.;
RL "The Arabidopsis GAI gene defines a signaling pathway that negatively
RT regulates gibberellin responses.";
RL Genes Dev. 11:3194-3205(1997).

```



RESULT 9  
ID 023643 PRELIMINARY: PRT: 532 AA.  
AC 023643.  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
DE RGA2 PROTEIN.  
GN RGA2.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
NCBI\_TaxID=3702;  
RX MEDLINE=97379310; PubMed=9237632;  
RA Truong H.N., Caboche M., Daniel-Vedele F.;  
RT "Sequence and characterization of two Arabidopsis thaliana cDNAs  
RT isolated by functional complementation of a yeast glr3 gdh1 mutant.";  
RL FEBS Lett 410:213-218(1997).  
DR EMBL; Y11337; CAA72178.1;  
DR Mendel; 24146; Arath;3051;24146.  
SQ SEQUENCE 532 AA; 58789 MW; 1AC719CA8072239 CRC64;

Query Match 49.8%; Score 1590; DB 10; Length 532;  
Best Local Similarity 54.2%; Pred. No. 4.3e-95;  
Matches 331; Conservative 72; Mismatches 106; Indels 102; Gaps 11;

QY 22 EDKMWVSAAGEEVEDELLAALGYKVRASDADVAOKLEOLEMAMGCVGAGAAPDS 81  
DB 11 DKKTMMHEEDDGNDELAVLGKVRSSSEADVAOKLEOLEMAMGCVGAGAAPDS 63  
QY 82 FATHLATDTVHNPDTLSSWESMLSELNAPPPLPPAPQLNASTSTVTGSGGYFDLPP 141  
DB 64 -LSOLATEGVHNPALTYLWDSMLTDLNPP-----SNQVEDD- 93  
QY 142 SVDSSTSYALRPTSPAGATAPADLSVDVDPKRMRTGSGTSSSSSSSLGCG--- 198  
DB 94 ---SSNAEDLKAIP-----GDALL-----NQFADISASSSSNOGGGDT 129  
QY 199 -----ARSSVEAAPPPVAAAANATPALPVVVDTQEGIRLVHALLACAEVQOENL 250  
DB 130 YTTNKRKLCNSVVE-----TTTATAESTRHVYLVDSQENGRVYHALLACAEVQOENL 184  
QY 251 SAAEALVKQIPLLAASQSGAMRKVAAYFGELARVFRFPQPDSSLLDAFADLLAHF 310  
DB 185 TVAEALVKGIGFLAVSQIGAMRQVATYFAEALARIYRL--SPQSPIDHSLSDTLQMHF 242  
QY 311 YESCYLKFANHTANQALILEAFAGCRVHVVDGIGKQGMWPALLQALALRGGPSPRL 370  
DB 243 YETCYLKFANHTANQALILEAFQGRKRVHVIDFSMSOGIQMWPALLQALALRGGPSPRL 302  
QY 371 TGVGPPROPDETALQOVGMKLAQFAHTIRVDROYGIVAAATADLEPMLQDEGEDNE 430  
DB 303 TGIQPPAPNDPDYLEVQCKLAHLAEALHVEFEYRGVANTADLADASML---ELRPS- 357  
QY 431 EPEVAVNSVEFMRHLAQPGALEKVLCTVRAVRPRIVTVVEQANHNHSGFLDRFTESL 490  
DB 358 ELESAAVNSVEFLKHLRPAIDKVLGVNQIKPEIFTVVQESNNHSPFLDFSTESL 417  
QY 491 HYSYTMEDSLBGSGSGGSPSEVSSGAAAAAPAAAGTDQVMSVEYLLGRQICNVVACGAERT 550  
DB 418 HYSYTLFQSLLE-----VPSGQDKVMSVEYLLKQICNVVACGAPRRV 459  
QY 551 ERHETLGMWRNRLNAGFETVHLGNSAYKQASTLLALFAGGQGYVERKEGCTLGMWTR 610  
DB 460 ERHETLGMWRNRFSGAGFAAHIGSNARQKASMLALFNNGGQYVERESDGLGMWTR 519  
QY 611 PLIATSAWRLA 621

DB 520 PLIATSAWRLS 530

RESULT 10  
ID 09SRP9 PRELIMINARY: PRT: 547 AA.  
AC 09SRP9.  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
DE RGA1-LIKE PROTEIN.  
GN T21P5.13.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
NCBI\_TaxID=3702;  
RX MEDLINE=97379310; PubMed=9237632;  
RA Truong H.N., Caboche M., Daniel-Vedele F.;  
RT "Sequence and characterization of two Arabidopsis thaliana cDNAs  
RT isolated by functional complementation of a yeast glr3 gdh1 mutant.";  
RL FEBS Lett 410:213-218(1997).  
DR EMBL; Y11337; CAA72178.1;  
DR Mendel; 24146; Arath;3051;24146.  
SQ SEQUENCE 547 AA; 60493 MW; C4D18D5951D95634 CRC64;

Query Match 47.9%; Score 1530.5; DB 10; Length 547;  
Best Local Similarity 51.3%; Pred. No. 3.2e-91;  
Matches 325; Conservative 83; Mismatches 107; Indels 119; Gaps 14;

QY 17 GNGSSED---KMWVSAAGEEVE-----DELLAALGYKVRASDADVA 57  
DB 4 GGETWDDPPKPLPASRGSGSPMDKKADDNNNSNMDDLLAVLGKVRSSSEADVA 63  
QY 58 QKLEOLEMAMGCVGAGAAPDSFATHLATDTVHNPDTLSSWESMLSELNAPPPLPP 117  
DB 64 QKLEOLEMAMGCVGAGAAPDSFATHLATDTVHNPDTLSSWESMLSELNAPPPLPP 108  
QY 118 PAPQLNASTSTVTGSGGYFDLPPSVDSSTSYALRPTSPAGATAPADLSA----- 169  
DB 109 -----ASDLDLTRS-----CVDRSE--YDLRAIFG-----LSAFPKEEYF 143  
QY 170 DSVRPKRMRTGSGTSSSSSSSLGCGARSSVVEAAPPPVAAAANATPALPVVVVDQ 229  
DB 144 DEASKRIRLGSWCSSSDSESTRS-----VVLVDSQE 175  
QY 230 AGIRLVHALLACAEVQOENLSAAEALVKQIPLLAASQSGAMRKVAAYFGELARVFR- 288  
DB 176 TGVRLVHALVACAEIHOENLNLADALYKRVCTLGSQAGAMKVAITFAQALARIYRD 235  
QY 289 FRPQPD-SLLDAFADLLAHFESCPYLKFAHTANQALILEAFAGCRVHVVDGIGIKQ 347  
DB 236 TVAEADVCAAVNPSEFVELEMHFESCPYLKFAHTANQALILEAVTTARVAVIDGLNQ 295  
QY 348 GMQWPAALLQALALRGGPSPSPRLTGVGPPROPDETALQOVGMKLAQFAHTIRVDQYK 407  
DB 296 GMQWPAALLQALALRGGPSPSPRLTGVGPPROPDETALQOVGMKLAQFAHTIRVDQYK 355  
QY 408 VAATADLEPMLQDEGEDPNEPEVAVNSVEFMRHLAQPGALEKVLCTVRAVRPRIV 467  
DB 356 AAESLSDLEPEMF-----ETRPESETLVNSVFEFLHRLAAGSIEKLNTVKAIKPSI 409  
QY 468 VTVVEQANHNHSGFLDRFTESLHYSTMFDSLEGSSSGGSPSEVSSGAAAAAPAAAGTDQ 527  
DB 410 VTVVEQANHNHSGFLDRFTESLHYSTMFDSLEGSSSGGSPSEVSSGAAAAAPAAAGTDQ 453  
QY 528 VMSVEYLLGRQICNVVACGAERTERHETLGMWRNRLNAGFETVHLGNSAYKQASTLLAL 587  
DB 454 VMSVEYLLGRQICNVVACGAERTERHETLGMWRNRLNAGFETVHLGNSAYKQASTLLAL 513

```

QY 588 FAGGCGKYKEEGCCTTGCMTTPTLATSAMRLA 621
DB 514 YATGCGYRVEENDGCLMTGOTRPLTTISAMKLA 547

RESULT 11
Q9C8Y3 PRELIMINARY; PRT; 511 AA.
AC 09C8Y3;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE GIBBERELLIN REGULATORY PROTEIN, PUTATIVE.
GN T27F4.10.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehlner E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Corway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Eguu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultzer L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marzilli A.,
RA Milttscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzer S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Uterbeck T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsids
thaliana."
RL Nature 408:816-820(2000).
DR EMBL: AC020665; AAG52171.1;
SQ SEQUENCE 511 AA; 56754 MW; 1B60071697C92A9F CRC64;

Query Match 47.3%; Score 1511.5; DB 10; Length 511;
Best Local Similarity 52.3%; Pred. No. 4,9e-90;
Matches 324; Conservative 65; Mismatches 118; Indels 113; Gaps 13;

QY 1 MKREYODAGSGGGGGGSSDEKMYSAAGGEEVDELLAAGYKVRASMDADVAOKI 60
DB 1 MKREHHRRESSAGEG-SSSMYTVIKEBAAG---VDELLVYLGYKVRSSMDADVAHKI 54

QY 61 EOLEMAMGCVGAGAAPDDSFATHLATDTVHYNPTDLSWVESMSELNAPRPPPLPAP 120
DB 55 EOLEMYLGDG-----ISNLSDETHVYNPDSLGSWVESMISDLD--PTRIQEKP 100

QY 121 QLNASTSVTSGGYFDLPSPVSDSSSIYALRPIPSAGATAPADLSADSVDPKRMRT 180
DB 101 -----DSEYDLRAI--PGSAVYPRD--EHVYRKRRTKRI 130

QY 181 GGSSTSSSSSSSSSLGGGARSSVEAAPPVAAAANATPALPVVVVDVTQEGAIRLVHALLA 240
DB 131 ESRLSTRS-----VVLDLSQEGVRLVHALLA 158

QY 241 CAEAYOOENLSAAEALVKOIPLLAASOGAMRKVAAYFGEALARRFRRRPODSSILDA 300
DB 159 CAEAYOONNLKADALVKVHGLASSOAGAMRKVAAYFAGLARIRYRIYPRDVAL--S 216

QY 301 AFADLLAHFYESCPYLKFAHFTANQALIEAFAGCRRHVVDVGIRGQGMQWPLALAL 360
DB 301 AFADLLAHFYESCPYLKFAHFTANQALIEAFAGCRRHVVDVGIRGQGMQWPLALAL 360

```

```

DB 217 SFSDFLIHFYESCPYLKFAHFTANQALIEVFAETAEKVHYIDLGNHGLQWPLALQAL 276
QY 361 RRGPPPSFRLTGVGPQPDETDALQOVKMLAQFAHTIRVDYRGLVAATLADLEPMI 420
DB 277 RRGPPPSFRLTGVGYSLTP---IQEVGKMLGQLASTIGVNEFEKSIALNLSLDLPEMI 332

QY 421 QPGEEDPNEEPVIAVNVSEEMHRLLAQGLAEKVLGYTRAVRPRITVYVEQEAHNSG 480
DB 333 -----DIRGLESVAVNVSEFELHRLLAHGSIDKFLSTKSTRPITVYVEQEAHNSG 386

QY 481 TFIADRTESLHYSTMFDSLEGSSGSGPSEVSSGAAAPAAAGTDQVASEVYLRQIGN 540
DB 387 VFIDRTESLHYSSLSFDLSE-----GPPSQ-----DRVMSLELGRQILN 427

QY 541 VVACGAERTERHETVIGQNRNIGNAGFETVHIGSNAYVQASTLLALFAGGCGKYKEE 600
DB 428 LVACEEDREVERHETINOMNRNRFGLGFPVSGISNAYKQASMLALYAGADGVNVEEN 487

QY 601 GCLTLCMTTPTLATSAMRL 620
DB 488 GCLTLCMTTPTLATSAMRI 507

RESULT 12
Q65367 PRELIMINARY; PRT; 662 AA.
AC 065367;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE RGAL PROTEIN.
GN RGA-LIKE.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COL-0;
RA Sanchez-Fernandez R., Ardiles-Diaz W., van Montagu M., Inze D.,
RA May M.J.;
RT "Cloning of the Arabidopsis thaliana RGA-like gene, a putative member
of the YTH domain transcription factor family."
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ224957; CAI2242.1;
DR Mendel: 29006; Arabid.3051;29006.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 1.
DR SMART: SM00320; WD40; 1.
DR PROSITE: PS50082; WD_REPEATS_2; 1.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
DR Repeat; WD repeat.
SQ SEQUENCE 662 AA; 73126 MW; 958AB73D42121CA8 CRC64;

Query Match 46.8%; Score 1495.5; DB 10; Length 662;
Best Local Similarity 51.7%; Pred. No. 7,4e-89;
Matches 320; Conservative 68; Mismatches 118; Indels 113; Gaps 13;

QY 2 MKREYODAGSGGGGGGSSDEKMYSAAGGEEVDELLAAGYKVRASMDADVAOKI 61
DB 153 QREHNRRESSAGEG-SSSMYTVIKEBAAG---VDELLVYLGYKVRSSMDADVAHKI 206

QY 62 EOLEMAMGCVGAGAAPDDSFATHLATDTVHYNPTDLSWVESMSELNAPRPPPLPAP 121
DB 207 QLEMYLGDG-----ISNLSDETHVYNPDSLGSWVESMISDLD--PTRIQEKP 251

QY 122 LNASTSVTSGGYFDLPSPVSDSSSIYALRPIPSAGATAPADLSADSVDPKRMRT 181
DB 252 -----DSEYDLRAI--PGSAVYPRD-----EHVYRKRRTKRI 277

QY 182 GGSSTSSSSSSSSSLGGGARSSVEAAPPVAAAANATPALPVVVVDVTQEGAIRLVHALLA 241
DB 182 GGSSTSSSSSSSSSLGGGARSSVEAAPPVAAAANATPALPVVVVDVTQEGAIRLVHALLA 241

```



[illegible]

RESULT	13
Q9LF53	
ID	Q9LF53
AC	Q9LF53
DT	01-OCT-2000 (TREMBLrel. 15, Created)
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE	RGA-LIKE PROTEIN.
GN	K3M16-60.
OS	Arabidopsis thaliana (Mouse-ear cress).
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;	
LN	[1]
RP	SEQUENCE FROM N.A.
RA	Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Kotani H.,
RA	Tabata S., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.,
RL	submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
LN	[2]
RP	SEQUENCE FROM N.A.
RA	EU Arabidopsis research project;
RA	Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
EMBL	AL291150; CAC01893.1; -
Q9	SEQUENCE 523 AA; 57326 MW; 0F6CE0BD13403C35 CRC64;

Query Match	45.4%	Score 1450.5	DB 10	Length 523
Best Local Similarity	49.38%	Pred. No. 4.5e-86		
Matches 313	Conservative 76	Mismatches 115	Indels 131	Gaps 13

```

0Y      1 MKREYDA-----GGGGGGGSGSSEKKMVAAGGEYDELLAALG 46
      |||:::|-----|||:::|
Db      1 MKRSHQETSEEEAPSMVERKLENGGGG-----DDNM-----DEFLAVLG 42
      |||:::|-----|||:::|
0Y      47 KVRASDADVAQKLEQLEEMAGMGVGAGAPDPSFATLTATDVHYHPPTLSWVESML 106
      |||:::|-----|||:::|
Db      43 KVRSSDADVAQKLEQLEML-----SNDIASSNAFNDTVHNPEDLSGMAQSHL 93
      |||:::|-----|||:::|
0Y      107 SELNAPPPLPAPOLNASTSVTSGGGYFDLPSPVSDSSSLYALRPLPSPAGATAAPAD 166
      |||:::|-----|||:::|

```

[illegible]

RESULT	14			
09AS97		PRELIMINARY;	PRT;	493 AA.
ID	09AS97			
AC	09AS97;			
DT	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	PUTATIVE OSCAI.			
GN	PO707D10.30.			
OS	Oryza sativa (Rice).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;			
OC	Ehrhartoideae; Oryzae; Oryza.			
OX	NCBI_TaxID=4530;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV. NIPPONBARE;			
RA	Sasaki T., Matsumoto T., Yamamoto K.;			
RT	"Oryza sativa nipponbare(CA3) genomic DNA, chromosome 1, PAC			
RT	clone:PO707D10."			
RL	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AP002910; BAB40172.1; "			
QO	SEQUENCE 493 AA; 51864 MW; 5FDB670EB8899492 CRC64;			

Query Match	31.5%	Score 1005.5	DB 10	Length 493
Best Local Similarity	43.1%	Pred. No. 2.6e-57		
Matches 222; Conservative	64	Mismatches 136	Indels 93	Gaps 11

QY	113	PPPLPPAPQLNAHSSRYTGGGCFEDLP	PSYDSSSSLYALRPIPSGATAPADLSDV	172
Db	24	PPPAVAAD-----DVGITDPACADNAA-----	ALFEFAAPP-----	60
QY	173	RDPRMRTGGSSTSSSSSSSLGGRARSSV	EAPVAAAANATPALPVVVYDTQAGI	232
Db	61	-----CAPAATA-----	VLARREEEYAGI	82
QY	233	RLVHALLACAEAVO--OENLSAAEALVKQIPL	LAASOGGMKRVAAVFGELARVRFRR	290



```

Db      83 RLVLHMSGAGATFAGDHALASAOQLADSHALAAVSAASGIRVAHVETTLASRLFPSP 142
QY      291 PQPDSSLLDAAFADLLHAFVESCPLYKFAHFTANQATLEAAGCRVAVVPGIKQGMQ 350
Db      143 VAPPT--DAEHAFYH-HEFACPLYKFAHFTANQATLEAAGCRVAVVPGIKQGMQ 199
QY      351 WPAALQALALRPGPPSFLTGVPQPDDETDALQOVGKMLAOFATIRVDPQRLVAA 410
Db      200 WPAALQALALRPGPPSFLTGVPQPDDETDALQOVGKMLAOFATIRVDPQRLVAA 259
QY      411 TLADLEPMLQGEEDPNEPEVTAIVSVFEMHLLAOP--GALEKVLGTVRAVPRPI 467
Db      260 SLIDEPRMMLQ-----IAPGEAAVFNSTVLQHLRLGPDADQAPIDAVLDCVAVRPKI 312
QY      468 VTVVQGEANHNSGTFLDRTESLHYSTIMEDSLGSGSGGSEVSSGAAAPAAATIQO 527
Db      313 FTVVIDQEDHNTGTLDRTEALFYSAVEDSLD-----AASASGAGN 356
QY      528 VASEVYLGROIQNVVACGEAERTERHETLGOVRNLGNAGFEVHLGSAVYQOASTLLAL 587
Db      357 AMAEAYLQREICDVCGEGANRERHEPLSRNRDLTRAGLSAVPLGSNALQARMVLGL 416
QY      588 FAGDGYKVEKEGCLITLGMHTRPLIATSAMRLAG 622
Db      417 PS-GEHGSVEADGCLTLGMHGRPLFSASAMEAG 450

```

## RESULT 15

```

O9FUZ7 PRELIMINARY: PRT: 668 AA.
AC O9FUZ7:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE SCARECROW.
GN SCR.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20407145; PubMed=10948251;
RA Lim J., Helariutta Y., Specht C.D., Jung J., Sims L., Bruce W.B.,
RA Diehn S., Benfey P.N.;
RT "Molecular analysis of the SCARECROW gene in maize reveals a common
RT basis for radial patterning in diverse meristems."
RL Plant Cell 12:1307-1318(2000).
DR EMBL; AF263457; AAG13663.1;-.
SQ SEQUENCE 668 AA; 71162 MW; 92B7431046B2B621 CRC64;

```

Query Match 21.0%; Score 670; DB 10; Length 668;

Best Local Similarity 30.9%; Pred. No. 2e-35; Matches 202; Conservative 77; Mismatches 262; Indels 112; Gaps 19;

```

QY      8 AGSGGGGGGSGMSHDKMVS-----AAAGEGEVD-----ELLAALGYKVA 50
Db      79 AAAAGVGGSGAAPSASQALPRLTQLHQLPRAFCINHAREVDYRAHRAAANQAGGEATA 138
QY      51 SDMDVYAKLEQLEAMGNGVGAAGAPDDSPATHLADTVHYNPTDLSWVESML-SEL 109
Db      139 STTAVWDGIIRI---IGSSG--GAAVSTQLINHVREIIRPCNPGILASLELRLRSLD 192
QY      110 NAPPPLPAPQNLNSTSTVTSVSGGYFLPSPVSDSSSIYALRPSPAGATAPADLSA 169
Db      193 AADPAPLPPPO-----POHALLHGAPAAAPAGLTLPP-----PP 227
QY      170 DSVRPDKRMRTGSSSTSSSSSSSISGGARSSVVEAAPPVAAAANATPALPVV----- 224
Db      228 PLPLPKRRHHPPCQOQOQOQEEPHAPQSPKAPTAETAAAAAQAQAAAAAQAQAKERKEQ 287

```

```

QY      225 --VDQEGIRLVHALLACAEVQENLSAEALYKQIPLLAASOGAMRKVAAYPEAL 282
Db      288 RKRQREEGHLLTLLQCAEAVNADNLDADQOTLIELATPFGTSTQRAATRAEAM 347
QY      283 ARRVFR-----FRPDSSLLDAAFADLLHAFVESCPLYKFAHFTANQATLE 330
Db      348 SARLVSSCLIVAPLPFGS---PAAARLHGVAAMAFQVFNCGISPVKTSHPFANQAIGE 403
QY      331 ARAAGCRVAVVDPFGIKQGMQWPAALQALALRPGPPSFLTGVPQPDDETDALQOVGK 390
Db      404 APEREREVHIIDIDIMQIGQWPGLFHILASRPGPPRVRLTIGIA----SMALFATGKR 459
QY      391 LAQFAHTIRVDPQRLVAAATLADLEPMLQGEEDPNE---EEVYIAVSVFEMHRL 446
Db      460 LSDFADTLGLPFEFCAVA-----EKAGNVDPKLGTVRREAAVAVHMLH--HSL 505
QY      447 LAQPGALEKVLGTVRAVPRPIVVEQGEANHNSGTFLDRTESLHYSTIMEDSLGSGSG 506
Db      506 YDVTGSDSTIMLTIQRLAKKVITMVEQDLSH-SGSFLARFEALHYYSALFDSLDSYGE 564
QY      507 GGBSEVSSGAAAPAAAGTDQVNSEVYLGROIQNVVACGEAERTERHETLGOVRNLGNA 566
Db      565 DSPER-----HVEEQQLSREIRNVLAVGGPART-GDYKFGSWREKLAQS 608
QY      567 GFETVHLGSNAYKQASTLLALPAGDGYKVEKEGCLITLGMHTRPLIATSAMR 619
Db      609 GPPAASLAGSAAQAASLLIGMFP-SDGYTLVENGALKIGWKDCLITTSAMR 660

```

Search completed: January 22, 2002, 16:30:02

Job time: 337 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2002, 16:23:40 ; Search time 22.72 Seconds  
(without alignments)  
617.058 Million cell updates/sec

Title: US-09-485-529-7

Perfect score: 3194  
Sequence: 1 MKREYDADAGSGGGGGGMS.....TLGWHRLPLATSAWRLAGP 623

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, AA:\*  
1: /cgn2.6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2.6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2.6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2.6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2.6/ptodata/2/1aa/PCYUS.COMB.pep:\*  
6: /cgn2.6/ptodata/2/1aa/Backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query	Match Length	ID	Description
1	1594	49.9	532	US-09-117-853-2	Sequence 2, Appli
2	589.5	18.5	282	US-09-117-853-8	Sequence 8, Appli
3	569.5	17.8	259	US-09-117-853-6	Sequence 6, Appli
4	406.5	12.7	221	US-09-117-853-4	Sequence 4, Appli
5	155.5	4.9	166	US-09-117-853-10	Sequence 10, Appli
6	127.5	4.0	1891	US-08-804-227C-12	Sequence 12, Appli
7	127.5	4.0	1891	US-08-804-198-6	Sequence 6, Appli
8	122	3.8	1580	US-08-804-227C-11	Sequence 11, Appli
9	122	3.8	1580	US-08-804-198-5	Sequence 5, Appli
10	121	3.8	4130	US-09-428-517-2	Sequence 2, Appli
11	120.5	3.7	3816	US-08-922-635-22	Sequence 22, Appli
12	119.5	3.7	3816	US-09-428-517-3	Sequence 3, Appli
13	119	3.7	1384	US-08-976-255-11	Sequence 11, Appli
14	119	3.7	4545	US-08-804-227C-14	Sequence 14, Appli
15	119	3.7	4550	US-08-804-227C-8	Sequence 8, Appli
16	119	3.7	4550	US-08-804-198-2	Sequence 2, Appli
17	118.5	3.7	3567	US-07-642-734C-4	Sequence 4, Appli
18	118.5	3.7	3567	US-08-439-009A-4	Sequence 4, Appli
19	117.5	3.7	1719	US-08-459-568-4	Sequence 4, Appli
20	117.5	3.7	1719	US-08-399-411-4	Sequence 4, Appli
21	117.5	3.7	1719	US-08-516-859A-4	Sequence 4, Appli
22	117	3.7	5588	US-09-036-987A-6	Sequence 6, Appli
23	117	3.7	5588	US-09-370-700-6	Sequence 6, Appli
24	116.5	3.6	1093	US-08-545-860D-55	Sequence 55, Appli
25	116.5	3.6	1093	PCR-US94-04496-55	Sequence 55, Appli
26	116	3.6	486	US-08-821-355A-8	Sequence 8, Appli
27	116	3.6	486	US-09-003-687A-8	Sequence 8, Appli

28	116	3.6	486	US-09-136-605-8	Sequence 8, Appli
29	116	3.6	511	US-08-821-355A-9	Sequence 9, Appli
30	116	3.6	511	US-09-003-687A-9	Sequence 9, Appli
31	116	3.6	511	US-09-136-605-9	Sequence 9, Appli
32	115.5	3.6	4928	US-09-036-987A-5	Sequence 5, Appli
33	115.5	3.6	4928	US-09-370-700-5	Sequence 5, Appli
34	114	3.6	1298	US-08-690-473-2	Sequence 2, Appli
35	114	3.6	1298	US-09-259-821A-2	Sequence 2, Appli
36	114	3.6	1298	US-08-843-659-2	Sequence 2, Appli
37	113.5	3.6	829	US-09-413-814-105	Sequence 105, App
38	113	3.5	2441	US-08-194-468-2	Sequence 2, Appli
39	113	3.5	2441	US-08-961-739-2	Sequence 2, Appli
40	112	3.5	1706	US-08-459-568-2	Sequence 2, Appli
41	112	3.5	1706	US-08-399-411-2	Sequence 2, Appli
42	112	3.5	1706	US-08-516-859A-2	Sequence 2, Appli
43	111.5	3.5	1147	US-08-131-365B-38	Sequence 38, Appli
44	111.5	3.5	1147	US-08-668-123-38	Sequence 38, Appli
45	111.5	3.5	3729	US-08-804-227C-4	Sequence 4, Appli

#### ALIGNMENTS

RESULT 1  
US-09-117-853-2  
; Sequence 2, Application US/09117853

; Patent No. 6307126

; GENERAL INFORMATION:

Query Match 49.9%; Score 1594; DB 4; Length 532;  
Best Local Similarity 54.3%; Pred. No. 9, 7e-125;  
Matches 332; Conservative 71; Mismatches 106; Indels 102; Gaps 11;

QY	22	EDKMTYSAAGSEEDDELLAAGYKVRASDNADVAOKLEOLEMAMGSGVGAAPDD5	81
DB	11	DKTKMNEEDDNGDELLAAGYKVRSEMDVAKOLELVMM-----SNVEDD-63	
QY	82	FATHLTDTYVHYAPDLSWSVESMLSELNAPPPLPAPQALNASTSYTWSGGYDLP141	
DB	64	LSQALTEHYHYAPALYRWDSMLDLNPP-----93	
QY	142	SVDSSTIALRPIPSAPGATAPADLSADSVDPKRMRTGGSSSTSSSSSLGCG---198	
DB	94	SSNAEYDKAIP-----GDAITL-----NOFAIDSASSSSNOGGCGPT129	
QY	199	ARSSVYEAAPVAAANATPAPLVYVVDTOEAGTFLVHALACAAVOOENL250	
DB	130	YTTNKRKCSNGVE-----TTTATSESTRHYVLDVDSQENGVALHALACAAVOKENL184	
QY	251	SAAEALVKOIPLLAASQAGMKRVAAVFEALARVFRFPPODSSILDAFADLLHAHF310	
DB	185	TYAEALVKQIGFLAVSQIGAMKRVATYFAALARRIYRL--SPSQSPIDHSLSDTLOMHF242	

```
QY 311 YESCYLKAFAHNTANOALILEAFAGCRVHVVDGFKOGMOMFALLQALALRPGPPSRL 370
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 243 YETCYLKAFAHNTANOALILEAFAGCRVHVVIDESMSQGLQWALMALRGRGPPVRL 302
QY 371 TGVCPPODEDDALQOVGKLAQFAHTTRVDFQYRGVLAATLADLEPYMLDPEGEDPNE 430
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 303 TGIGPPADNFDYLFHEVCGKLAHLAEAIHVEFEYRGFAVNTLADLADASML---ELRPS- 357
QY 431 EPEVAVNSVFEHRLLOPGALEKVLGVRAVRPRITVYQEOANHNHSGTFLDFTSL 490
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 358 ELESVAVNSVFEHRLLOPGALIDVGLVNOIKPEFTVYQESNHNHSPITLDRFTSL 417
QY 491 HYYSTFDSLEGGSSGGSPSEVSSGAAAAPAAAGTDQVMSSEVYLGROIQNVVACGEART 550
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 418 HYYSTLFDLSLEG-----VPSGODKVMSEVYLGKQICNVVACGPPDRV 459
QY 551 EHEHETLGMRRNLGNAGFETVHLSGNATKQASTLALFAGGCGYVEKEGCLTIGMTR 610
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 460 EHEHETLSQWRNRFSGGAAAHIGSNARQASMLLALFNGGCGYVEESDCLMGWTR 519
QY 611 PLIATSAMRLA 621
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 520 PLIATSANKLS 530

RESULT 2
US-09-117-853-8
; Sequence 8, Application US/09117853
; Patent No. 6307126
; GENERAL INFORMATION:
; APPLICANT: Harberd, Nicholas P
; APPLICANT: Peng, Jinrong
; APPLICANT: Carol, Pierre
; APPLICANT: Richards, Donald E
; TITLE OF INVENTION: Nucleic acid encoding GAI gene of Arabidopsis thaliana
; FILE REFERENCE: 620-45
; CURRENT APPLICATION NUMBER: US/09/117, 853
; EARLIER FILING DATE: 1998-08-12
; EARLIER APPLICATION NUMBER: PCT/GB97/00390
; EARLIER FILING DATE: 1997-02-12
; EARLIER APPLICATION NUMBER: GB 9602796. 6
; EARLIER FILING DATE: 1996-02-12
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 8
; LENGTH: 282
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-117-853-8

Query Match 18.5%; Score 589.5; DB 4; Length 282;
Best Local Similarity 45.1%; Pred. No. 2.1e-41;
Matches 138; Conservative 35; Mismatches 54; Indels 79; Gaps 8;

QY 55 DVAQKLEOLEMAMGKGAGCAAPDSEFATHLATDTVHYNPTLSSWVESMLSELNAPP 114
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 27 DVAQKLEOLEVM-----SNVQEDD--LSQLATETVHYNPAELTYTWLDSMLTDLNP-- 76
QY 115 PLPPAPQLMASTSTVTGSGGYFDLPPSYDSSSTIYALRPISPAGATAPADLSADSVRD 174
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 77 -----SSNAEYDKAIP-----GDAIL- 93
QY 175 PKRMRTGGSSTSSSSSSSLGG-----ARSSVVEAAPPVAAANATPALPV 223
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 94 -----NQAIDSASSSSNOGGGDYTTNKRKLCSSNGVE-----TTTATAESTRHYV 140
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 224 VVDTOEAGIRLVHALLACAEAVQENLSAAEALVQIPLLAASOGAMRKVAAYFGEALA 283
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 141 LVDSENGVRVLHALLACAEAVQENLVAEALVQIGFLAVSQIGAMRKVATYFAELA 200
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 264 RRVFRFPQDPSLLDAAFADLLHAHFYESCPLYKFAHFTANOALILEAFAGCRVHVVD 343
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Db 201 RRIYRL--SPSOSPIDHSLDPTLQMHFETCPYKLFHFTANOALILEAFQGRKRVHVD 258
QY 344 GIKGM 349
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 259 SKSQGL 264

RESULT 3
US-09-117-853-6
; Sequence 6, Application US/09117853
; Patent No. 6307126
; GENERAL INFORMATION:
; APPLICANT: Harberd, Nicholas P
; APPLICANT: Peng, Jinrong
; APPLICANT: Carol, Pierre
; APPLICANT: Richards, Donald E
; TITLE OF INVENTION: Nucleic acid encoding GAI gene of Arabidopsis thaliana
; FILE REFERENCE: 620-45
; CURRENT APPLICATION NUMBER: US/09/117, 853
; EARLIER FILING DATE: 1998-08-12
; EARLIER APPLICATION NUMBER: PCT/GB97/00390
; EARLIER FILING DATE: 1997-02-12
; EARLIER APPLICATION NUMBER: GB 9602796. 6
; EARLIER FILING DATE: 1996-02-12
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 6
; LENGTH: 259
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-117-853-6

Query Match 17.8%; Score 569.5; DB 4; Length 259;
Best Local Similarity 45.2%; Pred. No. 8.6e-40;
Matches 135; Conservative 33; Mismatches 52; Indels 79; Gaps 8;

QY 55 DVAQKLEOLEMAMGKGAGCAAPDSEFATHLATDTVHYNPTLSSWVESMLSELNAPP 114
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 27 DVAQKLEOLEVM-----SNVQEDD--LSQLATETVHYNPAELTYTWLDSMLTDLNP-- 76
QY 115 PLPPAPQLMASTSTVTGSGGYFDLPPSYDSSSTIYALRPISPAGATAPADLSADSVRD 174
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 77 -----SSNAEYDKAIP-----GDAIL- 93
QY 175 PKRMRTGGSSTSSSSSSSLGG-----ARSSVVEAAPPVAAANATPALPV 223
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 94 -----NQAIDSASSSSNOGGGDYTTNKRKLCSSNGVE-----TTTATAESTRHYV 140
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 224 VVDTOEAGIRLVHALLACAEAVQENLSAAEALVQIPLLAASOGAMRKVAAYFGEALA 283
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 141 LVDSENGVRVLHALLACAEAVQENLVAEALVQIGFLAVSQIGAMRKVATYFAELA 200
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 264 RRVFRFPQDPSLLDAAFADLLHAHFYESCPLYKFAHFTANOALILEAFQGRKRVHVD 342
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 201 RRIYRL--SPSOSPIDHSLDPTLQMHFETCPYKLFHFTANOALILEAFQGRKRVHVD 257

RESULT 4
US-09-117-853-4
; Sequence 4, Application US/09117853
; Patent No. 6307126
; GENERAL INFORMATION:
; APPLICANT: Harberd, Nicholas P
; APPLICANT: Peng, Jinrong
; APPLICANT: Carol, Pierre
; APPLICANT: Richards, Donald E
; TITLE OF INVENTION: Nucleic acid encoding GAI gene of Arabidopsis thaliana
; FILE REFERENCE: 620-45
; CURRENT APPLICATION NUMBER: US/09/117, 853
; EARLIER FILING DATE: 1998-08-12
; EARLIER APPLICATION NUMBER: PCT/GB97/00390
; EARLIER FILING DATE: 1997-02-12
```

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

Query Match	12.7%;	Score 406.5;	DB 4;	Length 221;
Best Local Similarity	40.3%;	Pred. No. 2.7e-26;		
Matches 106; Conservative	30;	Mismatches 48;	Indels 79;	Gaps 8;

[illegible]

```

RESULT      5
US-09-117-853-10
; Sequence 10, Application US/09117853
; Patent No. 6307126
; GENERAL INFORMATION:
; APPLICANT: Harberd, Nicholas P
; APPLICANT: Peng, Jinrong
; APPLICANT: Carol, Pierre
; APPLICANT: Richards, Donald E
; TITLE OF INVENTION: Nucleic acid encoding GAI gene of Arabidopsis thaliana
; FILE REFERENCE: 620-45
; CURRENT APPLICATION NUMBER: US/09/117,853
; CURRENT FILING DATE: 1998-08-12
; EARLIER APPLICATION NUMBER: PCT/GS97/00390
; EARLIER FILING DATE: 1997-02-12
; EARLIER APPLICATION NUMBER: GB 9602796.6
; EARLIER FILING DATE: 1996-02-12
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 166
; TYPE: PR1
; ORGANISM: Arabidopsis thaliana
; US-09-117-853-10

```

Query Match	4.98;	Score 155.5;	DB 4;	Length 166;
Best Local Similarity	25.78;	Pred. No. 1.5e-05;		
Matches 61; Conservative	27;	Mismatches 50;	Indels 99;	Gaps 10

```

0Y 55 DVAKRLBDELEAKMGMGVGAAPDDSEFAHILHTDTHYNPLDLSNWSVEMLSTLNAPP 114
    |||||:::|
Db 27 DVAKRLBDELEVMM-----SNVQEDD--LSGLATETIYHNPAELTYWDSMLTDLPNPP-- 76
    |||||:::|

0Y 115 PLPAPQMLNASTSTVTGSGGYFDLPSPVSSSSITVLRPIPSFAGATAAPADUSADSVRD 174
    |||||:::|
Db 77 -----SSNNEVYLKAIPL-----GDAIIL- 93
    |||||:::|

```

OY	175	PKRRRTGGSSSTSSSSSSSGGARSSEVAPPAVAANAATPALVVVDPEAGIRL	234
Db	94	NQFALDASASSNSGGSG-----DYTTMKR-	118
OY	235	VHALLACAEAVQENLSAAELVYQIPLLAASOGGAMRKVAAYFGEA-LARRVYRFR	230
Db	119	---LKSNGVVEVTTPAAS-TTHVSMLT-----RRRYCVSTRRLWLKLERRR	165

RESULT 6  
US-08-804-227C-12  
Sequence 12, Application US/08804227C

```

1  GENERAL INFORMATION:
2  APPLICANT: Dehoff, Bradley S.
3  APPLICANT: Kuhstoss, Stuart A.
4  APPLICANT: Kosteck, Paul R., Jr.
5  APPLICANT: Sutton, Kimberly L.
6  TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
7  NUMBER OF SEQUENCES: 15
8  CORRESPONDENCE ADDRESS:
9  ADDRESSEE: THOMAS G. PLANT 1501
10 STREET: LILLY CORPORATE CENTER
11 CITY: INDIANAPOLIS
12 STATE: IN
13 COUNTRY: USA
14 ZIP: 46285
15
16 COMPUTER READABLE FORM:
17 MEDIUM TYPE: Floppy disk
18 COMPUTER: IBM Compatible
19 OPERATING SYSTEM: MS-DOS
20 SOFTWARE: ASCI(DOS) Text only
21 CURRENT APPLICATION DATA:
22 APPLICATION NUMBER: US/08/804,227C
23 FILING DATE: February 21, 1997
24 CLASSIFICATION: 435
25 ATTORNEY/AGENT INFORMATION:
26 NAME: Plant, Thomas, G.
27 REGISTRATION NUMBER: 35,784
28 REFERENCE/DOCKET NUMBER: X-8231
29 TELECOMMUNICATION INFORMATION:
30 TELEPHONE: 317-276-2459
31 INFORMATION FOR SEQ ID NO: 12:
32 SEQUENCE CHARACTERISTICS:
33 LENGTH: 1891 amino acids
34 TYPE: amino acid
35 TOPOLOGY: unknown
36 MOLECULE TYPE: peptide
37
38 US-08-804-227C-12

```

Query Match	4.08;	Score 127.5;	DB 2;	Length 1891;
Best Local Similarity	21.88;	Pred. No. 0.12;	Indels 189;	Gaps 31;
Matches 154;	Conservative 72;	Mismatches 290;		

QY	30	AAGGGEEDDELLAALGKRVRSAD---	NADVQAQKLEDELMAMGCV-----	GAGAPD	79
Db	347	ALGDTEIGALLATYTG--RERVGDPLMTLGSLSKNSIGHAQAAGVGVATIKVQAMHNSLP-	404		
QY	80	DSFATHLATDVAHN-PTDLSNWE---	SMLSLNAPPPLPPAPQDANASTSVTSGGG	135	
Db	405	-----RLTHVDAPSSKYMMSGAVELLTEGRSMNPRVERVRRRA-AVSASFVSGTNA	454		
QY	136	YFDLPYSDSSSYIALRPIS-PAGATAPAD--LSADSVADPKRMKTGSSSTSSSSSS	192		
Db	455	HVYLEEAPEVEGSHGGDGPEDRDATVGPLWVLSARS-REALRGAGRLAALARGRT	513		
QY	193	SSLGGARSSVVEAA-----	PPAAANANATPALPVPV	223	
Db	514	EGTGGGS-GLVVPADADICYSIATTRRETLERHAVALVOENRAGEDLAALAAAGRPESVYT	572		
QY	224	VVDIQEAGIRLVHALLCAEAVQOENTSAAEALVKQIPLLAASGGAMRKVAALFGEALA	283		

```

Db 573 GVARRGRI-----AFICSGGGAQR--LGAGRELGRPPVFA-----DALDEIAAEFDAHLE 622
QY 284 RR-----VERFRPODSSILD-----AAFA-----DLHAHFYESCPT 316
Db 623 RPLLSVFAEPATPDALDLDTDTYTOPALFAVETALFRLESQGLVPDVLVGH---SIGG 679
QY 317 LKFAH-----FTANQAILLEAFAGCRVHVVDGIGQMGWPAALLQALALRP-----362
Db 680 LVAHVAGVFSAADARLVASARGRLMRALPEGMAAAVQATEREAAALEPVAAGAVYAA 739
QY 363 -GGPSPFRITGVGPPODETDALQOVGWKLAQFAHTTIVDFQYRGVLAATLADLEPFMLQ 421
Db 740 VNGPQALVLSG-----DEAAVLAAGELARGRTKR-----771
QY 422 PEGEDPNEEPEVIAVNSVFEMHRLAOPGALKEVLTGVRAPRPIVTVVEQEAHNHSGT 481
Db 772 -----LRVSHAFISPRMDAMLADFRAVADTVDHARLPVSEV-----TGD 813
QY 482 FLDR-----FTESLHYSTMFDSLEGSSG-----GPSEVSSGAAAPAAAGTD 526
Db 814 LADAQQLTDPGYWTRQVQRPVRFADAVRTASARDAATFIELGPDVILCGMAEESLAAD 873
QY 527 QVMS-EVYLGROIQCNVACEGAERTERHETLQWRNRLGNAGFETVHLSNAYKO-----580
Db 874 VVFAPALRGREGDVTLRAAASAVVRGAGL-DWAALYGGTGARTTDLPTYAFQHSRYWL 932
QY 581 --ASTLLALFAGGDGDK---VEKEGCLTLGWHTRPLIATSAMRL 620
Db 933 APASAAVAPATPAAPSVRSVPFAEDQDGLMAAVHAGD-VASAAARL 976

```

## RESULT 7

US-08-804-198-6  
Sequence 6, Application US/08804198  
Patent No. 5945320

## GENERAL INFORMATION:

APPLICANT: Burgett, Stanley G.  
APPLICANT: Kustoss, Stuart A.  
APPLICANT: Rao, Nagaraja R.  
APPLICANT: Richardson, Mark A.  
APPLICANT: Rosteck, Paul R., Jr.  
TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PAUL R. CANTRELL 1138  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Macintosh  
OPERATING SYSTEM: Macintosh 7.0  
SOFTWARE: Microsoft Word 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,198  
FILING DATE:

## CLASSIFICATION: 435

## ATTORNEY/AGENT INFORMATION:

NAME: CANTRELL, PAUL R.  
REGISTRATION NUMBER: 36,470  
REFERENCE/DOCKET NUMBER: P9113  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-3885  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1891 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-804-198-6

Query Match 4.0%; Score 127.5; DB 2; Length 1891;  
Best Local Similarity 21.8%; Pred. No. 0.12;  
Matches 154; Conservative 72; Mismatches 290; Indels 189; Gaps 31;

```

QY 30 AAGECEYDELLAALGKVRSD---MADVAKLEQLEMMGMGCV-----GAGAPD 79
Db 347 ALGPIEGALLATYG--KERGDPMLGSLKSNIGHAQAAGGVKVVQAMRHGSLP- 404
QY 80 DSFATHLATDVHYN-PTDLSWYE---SMLEINAPPPLPPPOLNASTSTVTGSGG 135
Db 405 -----KTLHADABSKYEMASGAVELLTEGRSMRREVRERRA-AVSAFVGSTNA 454
QY 136 FYDLPPEYDSSSIYALRPDS-PAGATAPAD--LSADSVDPKRMRTGSGTSSSSSS 192
Db 455 HVYLEEAPVEAGSEHGDPGDPDPAVAGPLPVLISARS-REALRGQAGRIALAARQRT 513
QY 193 SSLGGARSSVVEAA-----PVVAAAANATPALPVV 223
Db 514 EGTGGGS-GLVVPADIGYSLATTRELEHRAVALVDENRTAGEDLALALAGRTPESVYT 572
QY 224 VVDIQEAGIRLVHALLACAEAVQENLSAAELVKQIPLLAASOGAMRKVAAVFGAALA 283
Db 573 GVARRGRI-----AFICSGGGAQR--LGAGRELGRPPVFA-----DALDEIAAEFDAHLE 622
QY 284 RR-----VERFRPODSSILD-----AAFA-----DLHAHFYESCPT 316
Db 623 RPLLSVFAEPATPDALDLDTDTYTOPALFAVETALFRLESQGLVPDVLVGH---SIGG 679
QY 317 LKFAH-----FTANQAILLEAFAGCRVHVVDGIGQMGWPAALLQALALRP-----362
Db 680 LVAHVAGVFSAADARLVASARGRLMRALPEGMAAAVQATEREAAALEPVAAGAVYAA 739
QY 363 -GGPSPFRITGVGPPODETDALQOVGWKLAQFAHTTIVDFQYRGVLAATLADLEPFMLQ 421
Db 740 VNGPQALVLSG-----DEAAVLAAGELARGRTKR-----771
QY 422 PEGEDPNEEPEVIAVNSVFEMHRLAOPGALKEVLTGVRAPRPIVTVVEQEAHNHSGT 481
Db 772 -----LRVSHAFISPRMDAMLADFRAVADTVDHARLPVSEV-----TGD 813
QY 482 FLDR-----FTESLHYSTMFDSLEGSSG-----GPSEVSSGAAAPAAAGTD 526
Db 814 LADAQQLTDPGYWTRQVQRPVRFADAVRTASARDAATFIELGPDVILCGMAEESLAAD 873
QY 527 QVMS-EVYLGROIQCNVACEGAERTERHETLQWRNRLGNAGFETVHLSNAYKO-----580
Db 874 VVFAPALRGREGDVTLRAAASAVVRGAGL-DWAALYGGTGARTTDLPTYAFQHSRYWL 932
QY 581 --ASTLLALFAGGDGDK---VEKEGCLTLGWHTRPLIATSAMRL 620
Db 933 APASAAVAPATPAAPSVRSVPFAEDQDGLMAAVHAGD-VASAAARL 976

```

## RESULT 8

US-08-804-227C-11  
Sequence 11, Application US/08804227C  
Patent No. 5876991

## GENERAL INFORMATION:

APPLICANT: Dehoff, Bradley S.  
APPLICANT: Kustoss, Stuart A.  
APPLICANT: Rosteck, Paul R., Jr.  
APPLICANT: Sutton, Kimberly L.  
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: THOMAS G. PLANT 1501  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285

```

: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: ASCII(DOS) Text only
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/804.227C
: FILING DATE: February 21, 1997
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Plant, Thomas, G.
: REGISTRATION NUMBER: 35,784
: REFERENCE/DOCKET NUMBER: X-8231
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 317-276-2459
: INFORMATION FOR SEQ ID NO: 11:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1580 amino acids
: TYPE: amino acid
: TOPOLOGY: unknown
: MOLECULE TYPE: peptide
: US-08-804-227C-11

```

```

Query Match          3.8%; Score 122; DB 2; Length 1580;
Best Local Similarity 20.7%; Pred. No. 0.26; Indels 208; Gaps 33;
Matches 135; Conservative 78; Mismatches 230;

```

```

OY 6 ODAGSGGGGGGMSSEDEKMMVSA--GGEVEDELLAAL 44
DB 301 QD-GASNGILTASGPAQOVIREALDAGLTPADYDAVEAHGTGTPIDPIAGALMATY 359
OY 45 GYKVASD---MADVAQKLEQLEMAMGMCV-----GAGAAPPDSEFATHLATDTVHYN 94
DB 360 GHE-RTGDPPLMGSLKSNIGHTQAAAGVAGVTKMYLALRHGELP-----RTLH-- 406
OY 95 PDDLSSWVE-----SMSELNAPPPPLPAPQOLNASTSTVTGSGGYE---DLPPSYDS 145
DB 407 ASTASSRIEMDAGAVELLDEAR-PMPRRAEGPRRAGISSFGISGTHNAHLVIEEPPA--- 462
OY 146 SSSIVALARIPIS---PAGATAPADLSADSVDRPKRMRTGSSSTSSSSSSSSSLOGGAR 200
DB 463 -----RPEPEAAQPPAPATTVPLISAAGAR-----SLREQARRLAHLAGHEE 506
OY 201 SSVVEAAPPVAAANATPRLPVVVYVDTQAGIRLVHALLACAEAVQOENLSAAEA----- 255
DB 507 ITAADAAARSAAITTRALSHRASVLDDBRA--LIDRLTALAEDRRKDPGVTVGEAGSGRP 563
OY 256 -----LVKQIPILASOGGAMRKVAAYFEBALARVFRFRP----- 291
DB 564 PVFVEPPGGSSQMTGMAEELLDRAVPVRAKAEBCARALAAHLDMSVL-DVLRDAPGAPPID 622
OY 292 -----QPPSSLLDAFAADLLHAH-----FYESCPLYKFAH----- 321
DB 623 RADVVQPTLFTMMVSLAALMESHGVRPAVVGHSOGCEITAAHAAAGLSIDDAARYAERS 682
OY 322 -----FTANQATILEPACGRVHVVDFTKQGMQ-WPALLQALALRPGGPPSFRLTGVP 375
DB 683 RLMKRLAGNGMLSYAPADRV-----RELMEPWAEMSMVAAY--NGPASTVAVG--- 730
OY 376 PQPDELDALQGV-----WKLA--QFA-HPIRVDFOYR-----GIVATLADL 415
DB 731 ----DARALEEGGRLSAAGVLRMPPLAGYDFGHSFOYE-QFRAELLDTLGTVRPTAARL 785
OY 416 EEPFML-----QPEGEEDPNEPEVIAVSVFEMHRLAQPGALEKVLGTVRAVFRIV 468
DB 786 -PEFSVTAAHEPEG-----LDAAY-WYRNRRREVEFASTLRTL--LREGHR 829
OY 469 TYVEDEANINSGTFLDRFTESLHYSTMFDSLGGSSGGGSEVSSGAFAA 519
DB 830 TEVEMGPHLLGAIDEVAEGVHATLATLHRGSGGIDRRSSVGAFA 880

```

```

RESULT 9
US-08-804-198-5
: Sequence 5, Application US/08804198
: Patent No. 5945320
: GENERAL INFORMATION:
: APPLICANT: Bargett, Stanley G.
: APPLICANT: Kuhstoss, Stuart A.
: APPLICANT: Rao, Nagaraja R. A.
: APPLICANT: Richardson, Mark A.
: APPLICANT: Rostek, Paul R., Jr.
: TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PAUL R. CANTRELL 1138
: STREET: LILLY CORPORATE CENTER
: CITY: INDIANAPOLIS
: STATE: IN
: COUNTRY: USA
: ZIP: 46285
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: Macintosh
: OPERATING SYSTEM: Macintosh 7.0
: SOFTWARE: Microsoft Word 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/804.198
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: CANTRELL, PAUL R.
: REGISTRATION NUMBER: 36,470
: REFERENCE/DOCKET NUMBER: P9113
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 317-276-3885
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1580 amino acids
: TYPE: amino acid
: TOPOLOGY: unknown
: MOLECULE TYPE: peptide
: US-08-804-198-5

```

```

Query Match          3.8%; Score 122; DB 2; Length 1580;
Best Local Similarity 20.7%; Pred. No. 0.26; Indels 208; Gaps 33;
Matches 135; Conservative 78; Mismatches 230;

```

```

OY 6 ODAGSGGGGGGMSSEDEKMMVSA--GGEVEDELLAAL 44
DB 301 QD-GASNGILTASGPAQOVIREALDAGLTPADYDAVEAHGTGTPIDPIAGALMATY 359
OY 45 GYKVASD---MADVAQKLEQLEMAMGMCV-----GAGAAPPDSEFATHLATDTVHYN 94
DB 360 GHE-RTGDPPLMGSLKSNIGHTQAAAGVAGVTKMYLALRHGELP-----RTLH-- 406
OY 95 PDDLSSWVE-----SMSELNAPPPPLPAPQOLNASTSTVTGSGGYE---DLPPSYDS 145
DB 407 ASTASSRIEMDAGAVELLDEAR-PMPRRAEGPRRAGISSFGISGTHNAHLVIEEPPA--- 462
OY 146 SSSIVALARIPIS---PAGATAPADLSADSVDRPKRMRTGSSSTSSSSSSSLOGGAR 200
DB 463 -----RPEPEAAQPPAPATTVPLISAAGAR-----SLREQARRLAHLAGHEE 506
OY 201 SSVVEAAPPVAAANATPRLPVVVYVDTQAGIRLVHALLACAEAVQOENLSAAEA----- 255
DB 507 ITAADAAARSAAITTRALSHRASVLDDBRA--LIDRLTALAEDRRKDPGVTVGEAGSGRP 563
OY 256 -----LVKQIPILASOGGAMRKVAAYFEBALARVFRFRP----- 291
DB 564 PVFVEPPGGSSQMTGMAEELLDRAVPVRAKAEBCARALAAHLDMSVL-DVLRDAPGAPPID 622
OY 292 -----QPPSSLLDAFAADLLHAH-----FYESCPLYKFAH----- 321

```

Dd	623	RADYVQPLFTMMVSLSAALMESHGVRPAAYVGHSGEGLAAHAACALSIDDAARYIAERS	682
Oy	322	-----FTANQAILEAFACRCRRHVVDFFGIKGMQ-WPALLQALRLPGCPSPFRLLTGWGP	375
Dd	683	RLMKRLAGNGMLSVMAADRV-----RELMEPMERMSVAIV--NCPASVTAG---	730
Oy	376	PQPDETDALOOGV-----WKLA--QFA-HITRPDFOYR-----GLVATTLADL	415
Dd	731	---DARALEEGGRLSAGVLRLWELAVDFEGSHPOVE-QFRAELLDTLGTVRPTAARL	785
Oy	416	EPPFML-----OPRGESDDPNPEFEVIIVNSVFEEHMRLLAQGALEKYGVRAVRPRIV	468
Dd	786	-PFESTVAAAHEPBG-----LDAAV-WYRMNRREVVESTILRTUL--LRGHR	829
Oy	469	TVEEQEAHNHGSTFLDRFTESLHYVSTMFDSEGGSSGGSPSEVGSAANA	519
Dd	830	TEVEKGPHPLIGALIDEVALAEGVNATHLATLTHRSSGGLDRFRSSVGAAFA	880
RESULT 10			
US-09-428-517-2			
Sequence 2, Application US/09428517			
Patent No. 6251636			
GENERAL INFORMATION:			
APPLICANT: Bellach, Mary C.			
APPLICANT: Shah, Sanjay Krishnakant			
APPLICANT: McDaniel, Robert			
APPLICANT: Tang, Li			
TITLE OR INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE			
FILE REFERENCE: 30062-20029.00			
CURRENT APPLICATION NUMBER: US/09/428.517			
CURRENT FILING DATE: 1999-10-28			
EARLIER APPLICATION NUMBER: 60/120.254			
EARLIER FILING DATE: 1999-02-16			
EARLIER APPLICATION NUMBER: 60/106.100			
EARLIER FILING DATE: 1998-10-29			
NUMBER OF SEQ ID NOS: 12			
SOFTWARE: PatentIn Ver. 2.1			
SEQ ID NO 2			
LENGTH: 4150			
TYPE: PRF			
ORGANISM: Artificial Sequence			
FEATURE:			
OTHER INFORMATION: Description of Artificial Sequence: Recombinant			
US-09-428-517-2			
Query Match 3.8%; Score 121; DB 4; Length 4150;			
Best Local Similarity 20.7%; Pred. No. 1.3;			
Matches 147; Conservative 69; Mismatches 209; Indels 286; Gaps 35;			
Oy	14	GCGGC-----GSSDDKM-----WYSAAA-----GEGEVDELIAL--GYKV	48
Dd	3199	GGGGVSYGLAEARTIMLDITYGGRVSYAAVNGPSSTVYSGGAQLADELLACCEREGVRA	3258
Oy	49	R-----ASDMADVOKLEOLEMAMGCMGVGAGARPDSFATHLTDTVNYHPRTLSSWV	102
Dd	3259	KRPVPDYASHSQMOQLDELLEAL-----ADTVRQDSSVPEFTVI-----ADMU	3304
Oy	103	ESMSELNAAPPPLPPAPOLNASTSSTVYGSGGYFDLPSPVSDSSSIYALRPISPAGAT	162
Dd	3305	D-----TTALDAGWYE-----	3315
Oy	163	APADLSADSVDPDKMRKTGGSSSTSSSSSSSLGGCAASSVYEAPPVAAAAATPALPV	222
Dd	3316	-----TNLRBTYRFQ-----BAYBELVDAQMGCAFECSPHPVLVPGITETL--	3356
Oy	223	VVVDTQEGIRLVNHLLACAFAVOOENLSAAEALVKOJPL--LAASGGAMRKV--AAFY	278
Dd	3357	---DFFDADAVALSLSRKDDEGLDFELTSLAEAFVQGVPDVMTHTAFEGGRPFVLYPLY-	3412
Oy	279	GEALARRRFRFPDSSILLDAADFADLLHANHFESCPLYLKFAHTANOAILAEAFAGCRRV	338

DB	3413	--AFQORIRWLHEEPLOQEPVDEAW	-----DAEFMS-----	3440
QY	339	HYVDFGIRKQGMQMPALLQALALRPGGPPSFRITGVGPPPOPDETALQOY	-----GWKLQ	393
DB	3441	-VVERG-----	-----DNTVSDLLSTDAEALHTLVLPALSSRRRR	3475
QY	394	FAHTTRVDFQYRGLVAANTLADLEPEMLQPEGEDBNEPEEYIAVNSVFEMHRIILAO	PGAL	453
DB	3476	VEHRRLOQDRYR-----	VEWKP-----PAALDEVLGGLFVVPRIGLADGVV	3519
QY	454	EVVLGTVRAVRIRITYVE-----	OEANHSG-----TFL-----DR-----	485
DB	3520	ARVVAALVTR-RGGEVSVVLELPTRPDRRAYAEVAVAGRGVSGVVSFLMDDRRSHSVVP	3578	
QY	486	--FTESLIHYSTIMFDSLEGSSGGSPS--	EVESGA-AAAPAAAGT--DOVMSEVY-----	533
DB	3579	ADLAASLVIALQALVLDL-----	GRYGEGRMLVYIRGANVAPSDAGVYIDVQAQVMQFCRV	3635
QY	534	LG-----	ROICNVACEGAERTERTHTTGLQWNRRLGNAGFETV	571
DB	3636	LGLEHPELMGLGLVLDLPVGDEVCRRFVGCVASAGREDQVAVRGSGVWRILVRA	---VV	3692
QY	572	HGSMNAYKQASTLLALFAGGDGYKVEKEGCUTLGHHTPLATSMWRLAG	622	
DB	3693	DGGGGGWPRGTVLV--TGGLG-----	GLGHN-----TARMLVG	3725
RESULT 11				
US-08-922-635-22				
Sequence 22, Application US/08922635A				
Patent No. 6033871				
GENERAL INFORMATION:				
APPLICANT: PILETZ, John E.				
TITLE OF INVENTION: DNA MOLECULES ENCODING IMITADLINE RECEPTIVE POLYPEPTIDES				
TITLE OF INVENTION: AND POLYPEPTIDES ENCODED THEREBY				
FILE REFERENCE: Corrected Sequence Listing				
Patent No. 6033871				
CURRENT APPLICATION NUMBER: US/08/922,635A				
CURRENT FILING DATE: 1997-09-03				
EARLIER APPLICATION NUMBER: 08/650,766				
EARLIER FILING DATE: 1996-05-20				
EARLIER APPLICATION NUMBER: 60/012,600				
EARLIER FILING DATE: 1996-03-01				
NUMBER OF SEQ ID NOS: 22				
SOFTWARE: PatentIn Ver. 2.0				
SEQ ID NO 22				
LENGTH: 1070				
TYPE: PRT				
ORGANISM: Homo sapiens				
US-08-922-635-22				
Query Match				
Best Local Similarity 3.8%; Score 120.5; DB 3; Length 1070;				
Matches 105; Conservative 67; Mismatches 220; Indels 133; Gaps 20				
QY	121	QUNASTSSIVT-----	GSgc-----YIDLPPSSDSSSIYALRIPPAAGATAPADIS	168
DB	566	QIRASIQDLKTYIAVIAKTPGTGSSPGGSFADGQPAERRASN--	DORQVEYAEELAPAPVE	623
QY	169	ADSVBDPKRMRTGSGTSSSSSSSLGGGARSVYEAPPYAAAAANATPALPVVVVDQ	228	
DB	624	VAPAPAAASASGPAKTPAPALASISALVPETPVDEAPAPPAEAPAOVPSSELIQATSE	683	
QY	229	E-----	AGIRLVHALLACAEVAQOENLSAAEALVQIPIILAS	266
DB	684	ENQIPSHLPACPSLRHVASLRGSAITIELPHSSIAEYENELRHLMSSVVFYQTPGLEVT	743	
QY	267	OGGAMRKVAATYGEALARKVYFRPQPDSSLLDAAPADLIHAFYES-CPIYKFAHF-TA	324	
DB	744	ACVLLSTRAVYR-VLHDLIRYFSEPIQDFWHQKNTDVNSSPFHLSQCFVLKSLDLSQV	801	



```

QY 325 NQALE---AFACGRNVHVDFFGKGMWRDIALALRPGGRSFRITGVGRPRDET 381
Db 802 NVGLFDQHFLLTGSTPMQVYTCITLRSYLTNCFGLQIMLVSS---LRRPSRPBDK 856
QY 382 DALQOVGSKLA-----QFANTIRVDEFOYRGIVAAATLADLERFMLOREGEEDRPNEBEVY 435
Db 857 DFYESEFNKTKTGKMENTLHSSRVKFTYPS--EEIGDLFTVAQKMAE--PEKAPALS 912
QY 436 AVNSVFEMHMLLQPGALEVLTGTVAAVRPIVTVVEQANINNSGFFDRFPESLHY-- 492
Db 913 ILTVQAFQVGGMPRGCGRGPL---RKITLLTSSEL---FLDDE--DCVHTPLP 959
QY 493 -----YST-----MFDLSLG-----GS 504
Db 960 EFAKEPQRDRYRIDDGRVRYRDLRVILMGVQYFQAALTGVFDVQCHDLMSGVTLDFGE 1011
QY 505 SGGGRSEVSSGANA-----APAAAGDQVMS-----EYLLRQI 538
Db 1020 VPGGPARASQREYQOVVFPVPSSESEKELITLAKQEWALCGREL 1064

```

```

RESULT 12
US-09-428-517-3
; Sequence 3, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniell, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029, .00
; CURRENT APPLICATION NUMBER: US/09/428, 517
; EARLIER FILING DATE: 1999-10-28
; EARLIER APPLICATION NUMBER: 60/120, 254
; EARLIER FILING DATE: 1999-02-16
; EARLIER APPLICATION NUMBER: 60/106, 100
; EARLIER FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 3
; LENGTH: 3816
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant
; US-09-428-517-3

```

Query Match	3.7%	Score 119.5;	DB 4;	Length 3816;		
Best Local Similarity	20.7%	Pred. No. 1.6;				
Matches 145;	Conservative	71;	Mismatches 207;	Indels 279;		
				Gaps 35		
QY	14	GGGGM- 	-GSSSEDKM- :::	-MYSAAA- 	-GGEEDVDELLAL-  :	-GKYV 48
Db	2251	GGGGWVSVGLSAERVRTMLDPTGGRVSVAAVNGPSTSVYSGDVGQALDELACERGRVA				2310
QY	49	R-----ASDMAADVAKTLEOLEMANGMGCVGGAAPDDSFATHLATIDVYHNYPT---				DLS 99
Db	2311	RRVPVYASHSQAOMQDLDELLEAL-----ADITPROHSSVPEFFSVTVDWMTDTTALDAG				2364
QY	100	SM-----	YESMLSE-LNA-----	PPPLPAPQ--LNA-SNSTVITGS-----		133
Db	2365	YWFNTLRQETVRFQEAEBGLVAGMGKAEVCECSFHPALVPEIQTLIDLADONAAVILGSLRQD				2424
QY	134	-GGYEDLPSS-----	VDSSSSITALRP--	IFS-----	PAGATAPADLSA	169
Db	2425	EGGDLRLLTSLAEAFVGVQPVWDTHAFEGEMPRFTVDLPYPRQOHYWKKRAPAR-----				2479
QY	170	DSVRDPKMRRTGGSSSTSSSSSSSLGGGANSVYEADAPVAAANATPALPVLVVVDTOE				229

Db	2480	-----GANTGDVAAGELTAAHPRLLGAVEMPDSDGLVL-TGQ	2516
Qy	230	AGIRLVHALLA-----CAEAQOENLSAAEALVKQIPLLAASQGGAM	271
Db	2517	ISLR-THRWLADHEVLGSVLLPCTAFVELVAQAADRAGVDVDELLELVARVLVLPDRGGIO	25757
Qy	272	RKVAAYGEALARRVFEFRPOPDSLLDAAFADLHANHFESCPLYLKEAHTANQATILEA	331
Db	2576	VRLLALGPSEADGRRSILQJLHSPREBA-----	2600
Qy	332	PAGCRVHVVDVGIKGOMWRPALLQALALRPGSPBSFKLT---GCGRPQDRTDALQOYG	388
Db	2601	-AGFNR-----WTRHASEFVV-PGQTGAARPTBPAGVWMPRPAAGAPVALASDR	2645
Qy	389	W-KLAQSAHNTIRDFQKGLVAA-----TLAD---LRFPMIQPGE---	425
Db	2646	YARLVEYRGVYTGSPSQ--GLHTAMRHGDVYAEVALPECTPAPDGYALHPALLDAVAQAVG	2703
Qy	426	-----EDPNE-----PEEYLAIVNSVFEMHHLNQPGAL	453
Db	2704	LGSEFVEPBGVVYLPFLMSDVTLLHATGATSLRWKSPAGDPVAL-----ALADPPAGA	2755
Qy	454	EKVLGVTAARVPRIVTVVEQEAHNHSGTFLDRTEESLHYSTMFDLSGGSS--G	506
Db	2756	P--VATVGAIRLRTTSAQ-----LARRGSAEHAHMERVEWEEGSSADRCRGA	2803
Qy	507	GGPSEVSSGAAPAAPAAAGTDQVMSEVYLHQIQINVAACGAEE	548
Db	2804	GGTTYEGERRAAEAGAAGTAAV-----LGRFVPAAYRTYKGV	2840

RESULT 13  
 US-08-976-255-11  
 Sequence 11, Application US/08976255  
 Patent No. 6136581  
 GENERAL INFORMATION:  
 APPLICANT: Jono, Keith E.  
 APPLICANT: Plowman, Gregory  
 TITLE OF INVENTION: KINASE GENES AND USES  
 NUMBER OF SEQUENCES: 53  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Lyon & Lyon  
 STREET: 633 West Fifth Street  
 STREET: Suite 4700  
 CITY: Los Angeles  
 STATE: California  
 COUNTRY: U.S.A.  
 ZIP: 90071-2066  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
 MEDIUM TYPE: storage  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: IBM P.C. DOS 5.0  
 SOFTWARE: Fastseq for Windows 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/976,255  
 FILING DATE: No. 6136581ember 21, 1997  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 60/031,675  
 FILING DATE: No. 6136581ember 22, 1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Walburg, Richard J.  
 REGISTRATION NUMBER: 32,327  
 REFERENCE/DOCKET NUMBER: 229/182  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (213) 489-1600  
 TELEFAX: (213) 955-0440  
 TELEX: 67-3510  
 INFORMATION FOR SEQ. ID NO.: 11:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1384 amino acids  
 TYPE: amino acid

STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-976-255-11

Query Match 3.7%; Score 119; DB 4; Length 1384;  
 Best Local Similarity 21.7%; Pred. No. 0.38;  
 Matches 120; Conservative 54; Mismatches 181; Indels 204; Gaps 28;

QY 116 LPAPQLNASTSVTSGGTFDLP-----PSYDSSSYIALRPI 155  
 | | | | | : : : : : | | | | | : : : : :  
 DB 764 LAPACVTPBTPWTASSCG--DHPQAEPLKATEAGTTGRPLRPLPSVPSQGA--PL 819  
 QY 156 PSPAGATAPADLADSVDRPMRTGSSSTSSSSSLGGARRSVVEAPVAAAN 215  
 | | | | | : : : : : | | | | | : : : : :  
 DB 820 PSEE-ASAPD--ADALPDSFTPTATGE--VSAIKLASALGSSSPVEVA--PSEDED 872  
 QY 216 ATPALPVVVDTQAGIRLVHALLACAAVOENLSAAALVKOI----- 260  
 | | | | | : : : : : | | | | | : : : : :  
 DB 873 TAATSGIFDTSSDGL-----QARRDVVPAFRSLQKQVGPDSLDLIDIPSSA 922  
 QY 261 -----PLAASOGGAMKV-AAIFGELARVFRFRQPPSSLIDA-----AFA 303  
 | | | | | : : : : : | | | | | : : : : :  
 DB 923 SDGGEVFPSPATGSPGGPRALDSGYDTEHY-----ESPEFVLKEAGCEGPQAF 974  
 QY 304 DLH-----AHFVESCPLYKFAHTANOAILLEAFAG---C--RRVHYVD 342  
 | | | | | : : : : : | | | | | : : : : :  
 DB 975 ELASEGEGPGETLSTLSGLNENKPTROSAVFSOLEAEKATSGPEKKCGDRAQPE 1034  
 QY 343 FGKIQNGMALLQALALRPG-----PSEFRLTGVG-----PPD 379  
 | | | | | : : : : : | | | | | : : : : :  
 DB 1035 LGLPSTGQ-PS-DQVCLRPGEVSGEAGSGPEVLPPLQEGSSPEPSTGSLVPEPP 1091  
 QY 380 ETDLDQY-----GKMLQFAITIRVDFOYKGLVAATLADLEPMLQPEGEDPNEPEVY 435  
 | | | | | : : : : : | | | | | : : : : :  
 DB 1092 EPQGPARYRPGSPSCQFVL-----LTPVPLRSGNSSEFGGP-- 1131  
 QY 436 AVNSVFEMHRLLAQALGALVIG-----TVRAVPRIVTVE--QEAHNSGFLDRFT 487  
 | | | | | : : : : : | | | | | : : : : :  
 DB 1132 -----GLSGPAPQKRMGCGPOTRAPRLALPGLPALLEGPEEEDSDSDSD 1182  
 QY 488 ESLHYSTMEDSLGSGSGGSPSEVS--SGAAAPRAAGTDQ-----VMSEVY 533  
 | | | | | : : : : : | | | | | : : : : :  
 DB 1183 BELCYSYQ-----EPSEDEEELAPVAVVVAESQARNLRSLKMPSLSETF 1231  
 QY 534 LGRQICNVACEGAERTER 552  
 | | | | | : : : : :  
 DB 1232 -----CEDLERKKK 1240

## RESULT 14

US-08-804-227C-14  
 ; Sequence 14, Application US/08804227C  
 ; Patent No. 5876991  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dehoff, Bradley S.  
 ; APPLICANT: Kuhloss, Stuart A.  
 ; APPLICANT: Rostock, Paul R., Jr.  
 ; APPLICANT: Sutton, Kimberly L.  
 ; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
 ; NUMBER OF SEQUENCES: 15  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: THOMAS G. PLANT 1501  
 ; STREET: LILLY CORPORATE CENTER  
 ; CITY: INDIANAPOLIS  
 ; STATE: IN  
 ; COUNTRY: USA  
 ; ZIP: 46285  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: MS-DOS

SOFTWARE: ASCII(DOS) Text only  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/804,227C  
 ; FILING DATE: February 21, 1997  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Plant, Thomas, G.  
 ; REGISTRATION NUMBER: 35,784  
 ; REFERENCE/DOCKET NUMBER: X-8231  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 317-276-2459  
 ; INFORMATION FOR SEQ ID NO: 14:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 4545 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: protein  
 US-08-804-227C-14

Query Match 3.7%; Score 119; DB 2; Length 4545;  
 Best Local Similarity 21.7%; Pred. No. 2.3;  
 Matches 143; Conservative 79; Mismatches 223; Indels 214; Gaps 35;

QY 9 GSGSGG---GGGM---GSSSEDKM-----MVSAAAGEGEE-----VDELLAA 43  
 | | | | | : : : : : | | | | | : : : : :  
 DB 1731 GGLMGLPYVGCGMMSVGAASEVYGVGEGLEWVSVAAVNGPRSVLSGDVGLSEYVAS 1790  
 QY 44 L---GYKVRASDMA-----DYAOKLEOLENAMKM-----GCVGAGA 76  
 | | | | | : : : : : | | | | | : : : : :  
 DB 1791 LMGDVEECRLDVSHGFHSLMEVLEGFRCGVESLEGRVPRGVVSVSGVGVSSGE 1850  
 QY 77 APDSEFATHLATDVFVHY-----NPTDLSWVE-----SMUSELANPPPLPAPQ 121  
 | | | | | : : : : : | | | | | : : : : :  
 DB 1851 LGDPGYWRARAREAVRFRADGVYGVGGLVGLVEGPRGVLTGAGELGAGDDVYVPA 1910  
 QY 122 L-----NASTSTV--TGS-GGYFDLPVSVDSSSYIALRPIPSP 158  
 | | | | | : : : : : | | | | | : : : : :  
 DB 1911 MRGRAREVEFEALATVFTRDAGLDATLHTGSTRIDL-PTYPQORDRYWLDPVFTA 1969  
 QY 159 AGATAPADLSDSVDRPMRTGSSSTSSSSSSSLGGARRSVVEAPVAAANATP 218  
 | | | | | : : : : : | | | | | : : : : :  
 DB 1970 VTGVEPAGSPADA-----RATERGRSTTA-----GIRYRNA-WQPAVVDKNGCP 2013  
 QY 219 ALPVVV---DTEAGIRLVHALLACAAVQOE-NLSAAALVKOIPLLAASOGAMRK 273  
 | | | | | : : : : : | | | | | : : : : :  
 DB 2014 AGHVLLAPDEDTDASGL-----APATARELAARGAEVHVAVPVGCREAGDGL 2064  
 QY 274 VAAYFGEALARRFRFRP-QPDSLLDAAFADLHAHFYESCPYLKFAHTANOAILLEAF 332  
 | | | | | : : : : : | | | | | : : : : :  
 DB 2065 RAAGDGAARSTRVLMAPAEPDAA--DAV-----ALVQALGEA- 2100  
 QY 333 AGCRVHVVDGFIKQGMQWPLLQALALRPGSPFRLTGVPQPPDETALQGVKMLA 392  
 | | | | | : : : : : | | | | | : : : : :  
 DB 2101 -----VEPAPLWITTRAAAVRPDETPS-----VGAGQL-----MGLG 2133  
 QY 393 QFAHTIRVDFOYRGLV---AATLADLEPFM--LOPEGEEDPNEPEVYIAVNSVFEMHRL 446  
 | | | | | : : : : : | | | | | : : : : :  
 DB 2134 QVA-ALEIGRRMGLADLPASPAVLRTFYGALLAGGENGFARPSVHRRVYP----- 2188  
 QY 447 LAQGALEKVLGTVRAVPRIVTVEQEAHNSGTFL--DRETESLHYSTMEDSLGEG 503  
 | | | | | : : : : : | | | | | : : : : :  
 DB 2189 -----APVPVPAASARVYTTAPATAVGEDARDSDVVPVDDRMSSG-----TVLLTG 2235  
 QY 504 SSGGSPSEVS-----SGAA-----AAPAAGTDQWSEV-YLGRQICNVACEGAER 549  
 | | | | | : : : : : | | | | | : : : : :  
 DB 2236 GTGALGAGVARRLARSGAARLLLVGRGAAGPVGELVEELTALGSEVA-VEACDVADR 2293

## RESULT 15

US-08-804-227C-8  
 ; Sequence 8, Application US/08804227C  
 ; Patent No. 5876991

```

GENERAL INFORMATION:
APPLICANT: Dehoff, Bradley S.
APPLICANT: Kubstoss, Stuart A.
APPLICANT: Rostock, Paul R., Jr.
APPLICANT: Sutton, Kimberly L.
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESSES:
ADDRESS: THOMAS G. PLANT 1501
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas, G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 4550 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-804-227C-8

```

Query Match 3.7%; Score 119; DB 2; Length 4550;  
 Best Local Similarity 21.7%; Pred. No. 2.3; Mismatches 223; Indels 214; Gaps 35;

```

Matches 143; Conservative 79; Mismatches 223; Indels 214; Gaps 35;
QY 9 GSGGG--GGGM--GSSDDKM-----WVSAAGEGEE-----VDELLAA 43
DB 1736 GGLMGGLPVGGGMSVGAESVRYGVEGIGEVSVAAVNGPRSVLSGDVGLSESVAS 1795
QY 44 L--GYKVRASDMA-----DVAOKLEOLEMANGM-----GGVGAGA 76
DB 1796 LMGDVECRRLDVSHGFHSVLMPEVLGEFGRVSELEFGVRPQVYVSGVGSGVSGSE 1855
QY 77 APDDSFATILATITVH-----NPTDLSWYE-----SKLSELNAPPPPLPPAPQ 121
DB 1856 LGDPGYVWRHAREAVRFADGVVGRGLGVTGVLVEVPHGVLTGMAGECIGAGDDVVVPPA 1915
QY 122 L-----NASTSSTV--TGS-GGYFDLPSSVDSSSIYALRPIPS 158
DB 1916 MRGRAREVEFEALATVFTTRDGLDATALHTGSTGRIDL-PTYPQORDRYWLDPVRTA 1974
QY 159 AGATAPADLSADSVDPKRMRTGSSSTSSSSSSSSSSIGGARSSVVEAAPVAAAANATP 218
DB 1975 VVGVEPAGSPADA-----RATERGRSTTA-----GIRYRVA-WQPAVVDGRNGP 2018
QY 219 ALPVVVV---DTQAGIRLVHALLACAEAVQD-NISAAEALVKQIPLLAASOGGAMRK 273
DB 2019 AGHVLLIAPDEPDADSGL-----APAIAELAVGAIEVHTVAVPGTGREAAAGDIL 2069
QY 274 VAAVFGAALARVYFRFRP-OPDSSILDAFAADLLHAHFYESCPLYKFAHFTANQAILAEF 332
DB 2070 RAAGDGAARSTRVLMALPAEDDA--DAY-----ALVOALGEA- 2105
QY 333 AGCRHVHVDFGIKQGMQWMPALLQALALRPGPPSFRITGVGPPODETALQOVGWKLA 392
DB 2106 -----VPEAPLMTTTRREAAVVRPDETPS-----VGGAQL-----WGLG 2138

```

```

QY 393 QFAHTIRVDFQYRGLV-----AATLADLEPEM--LOPEGEEDPNEEPEVIAVNSVFEMHRL 446
DB 2139 QVA-ALELGRRMGGLADLPSSASPAVLRTVYGALLAGGEQFAVRSGVRRVVP----- 2193
QY 447 LAQPGALEKVLGTVRAVRRPRTVTVVEQEAHNHNSGTF--DRTESLHYYSYMPDSIEGG 503
DB 2194 ----APVPVPASARTVTTAPATAVGEDARNDTSIDVVVPDDRWSG-----TVLITG 2240
QY 504 SSGGGRSEVS-----SGAA-----AAPAACTDQYMSFY-YLGRQICNVVACEGAER 549
DB 2241 GTGALGAQVARRLARSGAARLLVRRGAAGPGVGEIVELTALGSEVA-VEACDVADR 2298

```

Search completed: January 22, 2002, 16:24:19  
 Job time: 39 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 22, 2002, 23:03:40 ; Search time 3021.93 Seconds  
(without alignments)  
11600.701 Million cell updates/sec

Title: US-09-485-529-14

Perfect score: 2125  
Sequence: 1 atagagagcgcagtagctc.....tggatgagcagcactccg 2125

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba: \*  
2: gb\_htg: \*  
3: gb\_in: \*  
4: gb\_ov: \*  
5: gb\_om: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sy: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vl: \*  
15: gb\_da: \*  
16: em\_fun: \*  
17: em\_fun: \*  
18: em\_in: \*  
19: em\_om: \*  
20: em\_or: \*  
21: em\_ov: \*  
22: em\_pat: \*  
23: em\_ph: \*  
24: em\_pl: \*  
25: em\_ro: \*  
26: em\_sts: \*  
27: em\_sy: \*  
28: em\_un: \*  
29: em\_vl: \*  
30: em\_htgo\_hum: \*  
31: em\_htgo\_inv: \*  
32: em\_htgo\_rnd: \*  
33: em\_htg\_hum: \*  
34: em\_htg\_inv: \*  
35: em\_htg\_rnd: \*  
36: em\_htg\_other: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2125	100.0	2125	6	AX005805 Sequence
2	1956.2	92.1	2709	6	AX005794 Sequence
3	1870.4	88.0	1872	6	AX005794 Sequence
4	1502.4	70.7	1768	6	AX005804 Sequence
5	1379.8	64.9	1746	6	AX005848 Sequence
6	1341.4	63.1	2500	8	AB030956 Oryza sat
7	1341.4	63.1	122497	8	AC087797 Oryza sat
8	1317	62.0	2255	6	AX005806 Sequence
9	1309.4	61.6	1890	8	AX005806 Sequence
10	1220.2	29.2	800	8	AF377621 Zea mays
11	617	29.0	800	8	AF377621 Zea mays
12	617	29.0	800	8	AF377627 Zea mays
13	617	29.0	800	8	AF377627 Zea mays
14	617	29.0	800	8	AF377633 Zea mays
15	617	29.0	800	8	AF377633 Zea mays
16	615.4	29.0	800	8	AF377626 Zea mays
17	615.4	29.0	800	8	AF377631 Zea mays
18	615.4	29.0	800	8	AF377634 Zea mays
19	615.4	29.0	800	8	AF377635 Zea mays
20	615.4	29.0	800	8	AF377636 Zea mays
21	615.4	29.0	800	8	AF377639 Zea mays
22	615.4	29.0	800	8	AF377640 Zea mays
23	615.4	29.0	800	8	AF377642 Zea mays
24	615.4	29.0	800	8	AF377643 Zea mays
25	613.8	28.9	800	8	AF377622 Zea mays
26	613.8	28.9	800	8	AF377623 Zea mays
27	613.8	28.9	800	8	AF377625 Zea mays
28	613.8	28.9	800	8	AF377628 Zea mays
29	613.8	28.9	800	8	AF377629 Zea mays
30	613.8	28.9	800	8	AF377630 Zea mays
31	613.8	28.9	800	8	AF377632 Zea mays
32	613.8	28.9	800	8	AF377637 Zea mays
33	613.8	28.9	800	8	AF377638 Zea mays
34	613.8	28.9	800	8	AF377641 Zea mays
35	613.8	28.9	800	8	AF377644 Zea mays
36	420.4	19.8	770	6	AX005803 Sequence
37	382	18.0	132699	6	AC006917 Genomic s
38	381.8	18.0	1964	6	ATY15193 Arabidopsi
39	381.8	18.0	1964	6	ATY15193 Arabidopsi
40	380.2	17.9	1951	8	ATRG42 Y11337 A. thaliana
41	380	17.9	1779	6	AX081276 Sequence
42	378.4	17.8	1779	6	AX081278 Sequence
43	367	17.3	416	6	AX005809 Sequence
44	359.8	16.9	26604	8	ATK3M16 Arabidops
45	359	16.9	436	6	AX005863 Sequence

#### ALIGNMENTS

RESULT 1  
LOCUS AX005805 2125 bp DNA PAT 24-AUG-2000  
DEFINITION Sequence 14 from patent WO9909174.  
ACCESSION AX005805  
VERSION AX005805.1 GI:99288602

#### KEYWORDS

SOURCE

#### ORGANISM

bread wheat.  
Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Pooidae; Triticeae; Triticum.

1 (bases 1 to 2125)

REFERENCE Harberd, N.P. and Peng, J.

Genetic control of plant growth and development

Patent: WO 9909174-A 14 25-FEB-1999;

HARBERD NICHOLAS PAUL (GB); PENG JINRONG (GB)

LOCATION/Qualifiers

FEATURES

source

1..2125  
/organism="Triticum aestivum"



Db	2041	GAATTGACGACCGGCTACGATCATCGGGCTACGGGTTCCGCTCCGTCTGCGCTGA	2100
0Y	2101	agaggtagatgacgacgacgacccg	2125
Db	2101	AGAGGTGATGACGACGAACTCCG	2125
RESULT	2		
AX005794			
LOCUS	AX005794	2709 bp	DNA
DEFINITION	Sequence 3 from Patent WO9090174.		PAT
ACCESSION	AX005794		
VERSION	AX005794.1	GI:9928799	
KEYWORDS			
SOURCE			
ORGANISM			
		bread wheat.	
		Triticum aestivum	
		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
		Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;	
		Poideae; Triticeae; Triticum.	
REFERENCE		1 (bases 1 to 2709)	
AUTHORS		Harberd,N.P. and Peng,J.	
TITLE		Genetic control of plant growth and development	
JOURNAL		Patent: WO 9009174-A 3 25-FEB-1999;	
		HARBERD NICHOLAS PAUL (GB); PENG JINRONG (GB)	
FEATURES			
source		1..2709	
		location/Qualifiers	
		/organism="Triticum aestivum"	
		/db_xref="taxon:4565"	
		6	
		/note="n is any nucleotide"	
		11	
		/note="n is any nucleotide"	
		46	
		/note="n is any nucleotide"	
		85	
		/note="n is any nucleotide"	
		107	
		/note="n is any nucleotide"	
		407	
		/note="n is any nucleotide"	
		417	
		/note="n is any nucleotide"	
		430	
		/note="n is any nucleotide"	
		436	
		/note="n is any nucleotide"	
		448	
		/note="n is any nucleotide"	
		452	
		/note="n is any nucleotide"	
		490	
		/note="n is any nucleotide"	
		498	
		/note="n is any nucleotide"	
		536	
		/note="n is any nucleotide"	
		562	
		/note="n is any nucleotide"	
		727	
		/note="n is any nucleotide"	
		769	
		/note="n is any nucleotide"	
		1444	
		/note="n is any nucleotide"	
		1450	
		/note="n is any nucleotide"	
		1556	
		/note="n is any nucleotide"	
		1742	
		/note="n is any nucleotide"	
		1808	
		/note="n is any nucleotide"	

[illegible]

[illegible]

QY	1797	CGGGAAGCGCTACAAGTGTGGGAGGAAGAAAGCGTCGCTAGCGCTGGGGGTGGACAGCG	1856
Db	1925	GGGGACAGCGGCTACACAGTGGAGAAAGAAAGAGAGCGCTGCTGACGCTGGGGGTGGACACAC	1983
QY	1857	CCGCGTATCGCCACCTCGGCATGCGCGCTGCGCGCGCGCGTATCGCGAGTtttgac	1916
Db	1984	CCCCCTGATGTCACCTCGGCATGTGGCGCTGGCGCGCGCGTATCTCCGAGTTTGAAC	2043
QY	1917	gctgtaagtacacatcgtgagcatygaagaaacaacagcccgagcgcccgccgctct	1976
Db	2044	GCTGTAGTACACATCGTGAGCATGAGACACACACCCCGCGCGCGCCGCTCT	2103
QY	1977	ccgggaacgcacgacgcacgcacgttgaagaagaagaagctaaatgcatytcagttag	2036
Db	2104	CCGGGAACGACGACGACCGACCGACCTTGAAGAAGAAGAAAGCTAATGTCAATGAGTGAG	2163
QY	2037	cgcgtaatgcagagaccggtcagatcgatcggctcagggctcagggttcctcgtcgtcgc	2096
Db	2164	CGCTAATTTGCGACGCGACCGGCTACGATGATCGGGCTACGGGTGCTCCGCTCGCTGGC	2223
QY	2097	gtgaagaggtgagatgagacgacgaactcgc	2125
Db	2224	GTGAAGAGTGTGATGAGACGACGACTCCG	2252
RESULT	3		
LOCUS	TAE242531	1872 bp	DNA
DEFINITION	Triticum aestivum rht-D1a gene for gibberellin response modulator.		28-JUL-1999
ACCESSION	AJ242531		
VERSION	AJ242531.1	GI:5640156	
KEYWORDS	gibberellin response modulator; rht-D1a gene.		
SOURCE	bread wheat.		
ORGANISM	Triticum aestivum		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.		
AUTHORS	1 (bases 1 to 1872)		
	Peng,J., Richards,D.E., Hartley,N.M., Murphy,G.P., Devos,K.M., Flintham,J.E., Beales,P., Fisk,L.J., Worland,A.J., Pelica,F., Suchanek,D., Christou,P., Snape,J.W., Gale,M.D. and Harberd,N.P.		
TITLE	'Green revolution' genes encode mutant gibberellin response modulators		
JOURNAL	Nature 400 (6741), 256-261 (1999)		
MEDLINE	99347734		
REFERENCE	2 (bases 1 to 1872)		
AUTHORS	Harberd,N.P., Peng,J. and Richards,D.E.		
TITLE	Green revolution genes encode mutant gibberellin response modulators		
JOURNAL	unpublished		
REFERENCE	3 (bases 1 to 1872)		
AUTHORS	Richards,D.E.		
TITLE	Direct Submission		
JOURNAL	Submitted (25-MAY-1999) Richards D.E., Molecular Genetics, John Innes Centre, Colney Lane, NR4 7JQ, UNITED KINGDOM		
FEATURES	Location/Qualifiers		
source	1..1872		
	/organism="Triticum aestivum"		
	/db_xref="taxon:4565"		
gene	1..1872		
	/gene="rht-D1a"		
	/gene="rht-D1a"		
	/note="GAI ortholog"		
	/codon_start=1		
	/product="gibberellin response modulator"		
	/protein_id="CA8155.1"		
	/db_xref="GI:5640157"		
	/translation="MKREYDAGSGGGGCGGSSSEDKMVSAAAGEEVEDELLAALGYKRAADMDVAQKLEOLEWAMGMCVGAAPDDSFATHLADTHYNPDTLSSWVESMSEINAPPEPLPPAPOLNASTSTVGSGCFDLPFSSSSSYALRIPSPAGATPAVDVADSVDRPKEMRTGSGSTSSSSSSSLGGAGSSVEAPVAAANAPRALTPVDVDEAGIRIVAHILACFAVDQENFSAEFLVIOITPLAASOGAMKVA		
CDS			





[illegible]

QY	1342	tcagctcttgagatgatcacgagcttgctcgagcagcccgcgcccttgagaaggttcctgggc	1401
Db	721	TCAGTCTTCGAGAGACACCGGCTGCTGCGCACCCGGCGCCTTGAGAAAGGTCTCTGGGC	780
QY	1402	accgtgagcgacgtgagcgccagatctgtacagtgtgagagagagagcgatatacac	1461
Db	781	ACCGTGCCTCCGCGCCGAGATGCTACCGTGTGTGAGCAGAGAGCGAATCACAAC	840
QY	1462	tcgagcaattcttcgagacgcttaccagtgctctgcaactactccacaatgttcgat	1521
Db	841	TCCGGCACATTCGAGAACCGGCTTACCAGAGTCTCTGACTACTACTCCACCAATGTTCAT	900
QY	1522	tcctctgagagggcgagcagctccgagcgagcgccatccgaagtctcaatcgaggctctgct	1581
Db	901	TCCCTCGAGGGCGGCACTCCGGGGGGGCCATCCGAATCTCAATCGGGGCTCTCGTT	960
QY	1582	gtctctgacggcgcgcgagcagagcaagatcatctcgaagtgatcaacttgagcgagatc	1641
Db	961	GTCCTGCCCCCGCCGGCAGCGACAGTATCTCGAAGTGTACTTCGGCCGGCAGATTC	1020
QY	1642	tgcaacgtgtgagccttgcgagggggcgagcagacagcagcagagacgcctgggccaag	1701
Db	1021	TGCAACGTGTGGGCTCGAGAGGGGGCGAGCAGCAGAGGGCCACAGACGCTGGGGCAG	1080
QY	1702	tggcgagaaacgagctgggcaaacgcgggtcttgagaccgttcaacttggctccaatgctaac	1761
Db	1081	TGGCGGAACCGGCTGGGGCAACGCCGGGTTTGGAACCGTCCACCTGGGCTCCAAATGCTTAC	1140
QY	1762	aagcagcgagacagcagctgtgcgagctcttcgcgagcgagcgagcgtatacaaggttgagag	1821
Db	1141	AAGCAGCGGAGCACGCTGTGCGCTTTCGCCGGGGCGACGCGTACAAAGTGGAGAG	1200
QY	1822	aaggaagctgtcctgagcgttgagggtggcacaacgcgcgcgtgatacgccaactcgagatg	1881
Db	1201	AAGGAAGCTCTCTGAGCGTGGGGTGGCAACAGCGCCCGCTGATCCGCACCTTCGGCAATGG	1260
QY	1882	cgccctgagcgagcggtatctcgagagtttgaaagctgtaagtaacatgtgtgacatg	1941
Db	1261	CGCCTGGCGGGCGGTATCTCGGGAATTTTGAACCTGTAAATACACTCGTAGACATAG	1320
QY	1942	gaggaacaacagcccgcgcgcccgccgctctccgagcaacgacgacgacgacgac	2001
Db	1321	GAGGACACACAGCCCGGGGGCGCGCGGCTCTCGGGGAACGACGACGACGACGAC	1380
QY	2002	ttgaaagaagaagaaatcaatgtcatgtcatgtgagcgctgaaatcgacgacccgactacg	2061
Db	1381	TTGAAGAAGAAGAACTAAATGTCTATGTCTAGCGCTGAAATTGACGACGACCGGCTAG	1440
QY	2062	atcgatcgagctaaagggtgtgtccgtccgctcgtcgctgaaagaggtgatatgagcagcaac	2121
Db	1441	ATCGATCGGGCTACGGGTGTTCCGTCCGTCGTGCGCTGAAGAGGTGATGACGACGAAC	1500
QY	2122	tcgcg 2125	
Db	1501	TCCG 1504	
RESULT	5		
AX005848			
LOCUS	AX005848	1746 bp	DNA
DEFINITION	Sequence 57 from Patent WO990174.		
ACCESSION	AX005848		
VERSION	AX005848.1	GI:992843	
KEYWORDS			
SOURCE			
ORGANISM			
	bread wheat.		
	Triticum aestivum		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;		
	Poidea; Triticeae; Triticum.		
	1 (bases 1 to 1746)		
REFERENCE			
AUTHORS	Harberd, N.P. and Peng, J.		
TITLE	Genetic control of plant growth and development		





```

OY 1475 tggagcgtctacccgaggtctgctactactccaccatggttcgtatccctcggaggcg 1534
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 1681 tggacgggttaccgagctgctgctactactccaccatggttcgtatccctcggaggcg 1740
OY 1535 ggaagctccggcgcgccatccgaagtcatactcgaggagctgctgctcgtccggcg 1594
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 1741 ggagctc-----cgccacggccgagcctcttccgccgctccgctccg-----ggggcgggcg 1788
OY 1595 ccggacgagcagatgatactgctcgaagtgctacccctcgccgacagatctgcaacgtgagtg 1654
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 1789 gmggacgagcagcagatgctcgaagtgctacccctcgccgacagatctgcaacgtgagtg 1848
OY 1655 cctggaagggcgagagcgacagagcgccacagagctggtggcagtggtggaaacggcg 1714
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 1849 cgtgacgagggccggagccacgagcgcacgagcgcagcgtggggcagtgccgaacgcgcg 1908
OY 1715 tgggaacgcccgggttcgagacgctccactgtgctcactcaatgctcactcaagcagcgagca 1774
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 1909 tggcgccggccggcggcttgcagcccgctgacacacgctgacacacgctgacacacgca 1968
OY 1775 ggcctggtgctccttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1834
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 1969 ccctcctcgcgccttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2028
OY 1835 tgaagctgggggtgggcaacgcccgcgtgatgcacaccccgcatggtggcgtggcggcg 1894
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 2029 tgcagcgtggggcgagcagcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2088
OY 1895 cgtgactcgcgagtttg 1913
    ||||||| |||||||
DB 2089 cctgactcgcgaagtgttg 2107
    ||||||| |||||||

RESULT 7
LOCUS AC087797
DEFINITION Oryza sativa chromosome 3 BAC OSJNB0022E02 genomic sequence,
complete sequence.
ACCESSION AC087797
VERSION AC087797.5 GI:13699786
KEYWORDS HTG.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 122497)
Buell,C.R., Yuan,Q., Ouyang,S., Mofat,K.S., Hill,J.N.,
Gansberger,K., Brenner,M., Burgess,S., Hance,W., Shwartsbeyn,M.,
Tsirlin,T., Riggs,F., Hsiao,J., Zismann,V., Blunt,S., Pal,G.,
VanAken,S.E., Uteckback,T.R., Feldblum,T.V., Quackenbush,J.,
Salzberg,S.L., White,O. and Fraser,C.M.
Oryza sativa chromosome 3 BAC OSJNB0022E02 genomic sequence
Unpublished
2 (bases 1 to 122497)
Buell,R.
Direct Submission
Submitted (25-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
3 (bases 1 to 122497)
Buell,R.
Direct Submission
Submitted (20-APR-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
4 (bases 1 to 122497)
Buell,R.
Direct Submission
Submitted (03-MAY-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, rbuelli@tigr.org
On Apr 20, 2001 this sequence version replaced gi:13605985.
COMMENT Address all correspondence to:rice@tigr.org

```

## FEATURES

source

BAC clone OSJNB0022E02 is from *Oryza sativa* chromosome 3. The orientation of the sequence is from SP6 to T7 end of the BAC clone.

Genes were identified by a combination of several methods: Gene prediction programs including Fgenesh (<http://www.softberry.com/>), genscan and Genscan+ (Chris Burge, <http://CCR-081.mit.edu/GENSCAN.html>), GenemarkHMM (Mark Borodovsky, <http://genemark.biology.gatech.edu/Genemark/>), and Gensplicer (Minhela Ferreira and Steven Salzberg, contact [mpertea@tigr.org](mailto:mpertea@tigr.org)), searches of the complete sequence against a peptide database and the plant ESR database at TIGR (<http://www.tigr.org/tdb/tgi.shtml>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

Location/Qualifiers

1..122497

/organism="Oryza sativa"

/cultivar="Nipponbare"

/sub\_species="japonica"

/db\_xref="taxon:4530"

/chromosome="3"

/map="C944"

/clone="OSJNB0022E02"

/complement(1271..>1579)

/gene="OSJNB0022E02.11"

/complement(1271..1579)

/gene="OSJNB0022E02.11"

/note="predicted by fgenesh"

/complement(1271..1579)

/gene="OSJNB0022E02.11"

/codon\_start=1

/product="hypothetical protein"

/protein\_id="AAK50135.1"

/db\_xref="GI:13937304"

/translation="MENTKAAPLVSSPSVSLRAVDPRTLRAPPLIHGLSYAP  
SCTRSSGTLFVPTKLALMAKVYIDGWSHVSAPQKMAVLLENLIGSYFAVTS  
"

<1386..>1640

/gene="OSJNB0022E02.18"

1386..1640

/gene="OSJNB0022E02.18"

/note="Predicted by glimmer"

1386..1640

/gene="OSJNB0022E02.18"

/codon\_start=1

/product="hypothetical protein"

/protein\_id="AAK50133.1"

/db\_xref="GI:13937302"

/translation="MRASVFTGVSHNVPLERVROLGATLRPCSSGAGANNRGSTAA  
GSDNTEGDEDNDGAFFSITAPQNPQTFVATVREEL"

/complement(1010<5240..5517,5605..5684,6132..6212,  
6314..6383,6539..6613,6693..6898,7177..7426,7506..7551,  
8773..8842,9564..9633,11257..11341,11440..11524,  
11608..11698,12255..9633,11257..11341,11440..11524,  
11608..11698,12255..>12557)

/gene="OSJNB0022E02.2"

/complement(5240..12557)

/gene="OSJNB0022E02.2"

/note="EST D47656, AU088605 from this gene"

/complement(1010(5439..5517,5605..5684,6132..6212,  
6314..6383,6539..6613,6693..6898,7177..7426,7506..7551,  
8773..8842,9564..9633,11257..11341,11440..11524,  
11608..11698,12255..12388))

/gene="OSJNB0022E02.2"

/codon\_start=1

/product="unknown protein"

```
/protein_id="AAK50136.1"
/db_xref="GI:13937305"
/translation="MORRAQTWAGVGTAKTAAQAAAHAALFCEFTLLALKYDGRPTAYSM
WIRIPLMIFPHGIVARGFSEMPASLPGRHMAKCSIVAPLIIAFELILCTIESL
RVKSKPTVDLKIIVPLIPLAFVITLADNFRKRALMGDESKSDEAIWELPIFWPA
ISWVLIATFTTLKLISGDVGLGMDLFINIACFAFLVCTRMFNPIHSPNP
GEASSSSAARIYRDWESGLPLSLDEHQERLCPLIGGHVMIPLVIEVLLCMRL
EGPSPAOYIPDFWLFSPPLFLLGAGVLFSLARLLEKVVLLRNGPVPSPNLTISYK
RDCAPFLHRSGLGMMSIDGSKEOARLEFYSTGNTFCGYPRPEVVRKMPRDLA
EVRRLQALGGESEITKCTKOEFRRLQNEVLCRICYESEICWVLLPCRRITLCKTC
SDKKKCPICVPIIEERNPYDY"
repeat_region
complement(12328..12420)
/rpt_family="GAGC"
complement(12558..12580)
/rpt_family="GAG"
join(<14168..15023,15131..15274,15344..16687,16818..17803,
17887..18048,18262..18366)
/gene="OSJNB0022E02.12"
14168..18366
/gene="OSJNB0022E02.12"
/note="similar to gag pol polyprotein GB:AD12997
join(14168..15023,15131..15274,15344..16687,16818..17803,
17887..18048,18262..18366)
/gene="OSJNB0022E02.12"
/codon_start=1
/product="putative gag-pol polyprotein"
/protein_id="AAK50117.1"
/db_xref="GI:13937286"
/translation="MAGFADALRPDKFTGVHFKMQRVTLWLTAMKCFVSTKPEG
VLIAQOKQFEATTLFVGCTLSVGLVLYVYHMTDAKELMDALNTKFCATASND
LYIMEOPHDYKMAQNRVSEQAHIEIQTMAKLELLKVLPRKFAAGCIATLPSWRS
FGTALKRROEYVEGLIASLDVEKAREKAAGKGGSGSAVAVHAKQKSKGKYK
AOQTNRKROKNNNNPNODERTCFVCGOVHILARKCEPGRGAMVAPGOTSANVTI
GNTGDSGYGTGFHPRNGEWNTFCGSCMHGABSEYVFGKRAAEERAAECFYEOESC
YECGLFRPSLDFCCKNSVNTICGVDEAAMVHSLRCHINFGMLSRSSCLLPDS
IVKSKCHSCVQSKOPRPHKAEERNLAPLELHSDCEKNVGLTIGKRYEMLID
DATRCVYVLLKTDLEADLYFKYIKAEVNDLDRKIKRLRSDRGEEFSNFDLCEE
HGIIHERTPYSPESNGIAERKNRTLIDLVNAMDITAGLPKAMGEALITSNHLYNR
PNRKMDTPYEIWIWGRPSLSYLRMGCLAVNPITKRLGLKTVQCVLYAHS
IAYRFLIVKSEVPDMHYGTIMESRDATFESEFPFMKDHGSGNSGSPSETIPSETIPEQ
TEHHELVSEEDYSEAPRSKRQRTAKSFGDPLVYLDIPKISIAEYASPDADYRK
EAVSEMDSTIANGTWEYTERPYGCKPRGCVKWRKKIKRPASHGILVHONDVTAFL
NGELDEETIMDQPGFVVEGQEVCKLKLKSLYLKQAPQMHKFKTLTISAGFVN
EADKCYVYRHGEGGVILCYLVDDILJFTGMLVYINFKSLSONFMKDLGVADVIL
NIKLIRSENGITLLQSHYVEKILNRFYIDSKPPTPYDPELLRKNKRIARDNEYS
QITGLSYIASATRPDISFAVSKLSRFTSNPNDHMRALREVMYILGTVELGHYTG
YPAILEGYSDSNMTSDVDEIKATSGVYPTLGGGAVSMRSCQTLITSTMAELTALD
TATVBAEMLRDMLDLPVEKPTCEKTEIYVCOETIKRLKSTVNGLHPNSEKFRSLHES
TITVCDRQCIEGDGFETHSILEGTHLCELDCWSQSMKILGSSRKLTKDLGV"
repeat_region
complement(18822..18858)
/rpt_family="CAGGC"
complement(19612..19640)
/rpt_family="AT_rich"
complement(20459..20487)
/rpt_family="AT_rich"
<21271..>21576
/gene="OSJNB0022E02.13"
21271..21576
/gene="OSJNB0022E02.13"
/note="Predicted by tgenesh"
21271..21576
/gene="OSJNB0022E02.13"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAK50120.1"
/db_xref="GI:13937289"
/translation="MTTRTVLSEYSDSGHNSGRNAGVAVNGHGVAGGAGGLHGYI
ACAVLLASTPYALHIOGDELQGVMTAKODCWQRQAPANDVDYEPITSNDASVPSN"
complement(22507..22529)
/rpt_family="(TA)n"
complement(22544..22588)
/rpt_family="AT_rich"
```

```
repeat_region complement(23124..23217)
/rpt_family="(GCAGAA)n"
complement(join(<23165..23588,23765..23917,25379..>25551))
/gene="OSJNB0022E02.14"
complement(23165..23551)
/gene="OSJNB0022E02.14"
/note="Predicted by tgenesh"
complement(join(23165..23588,23765..23917,25379..25551))
/gene="OSJNB0022E02.14"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAK50123.1"
/db_xref="GI:13937292"
/translation="MAAGSGARAVDNGCGVGGPPAPASLPPIPRHSSPRRRTPA
SPGILPRLRCOYSTGSLPEBSAAGVAVAPRSPRSPRADADEGLGSSAAV
WMAQARSCEGNAESMAVLSACPYADGLALAGATHGCGKVGYHGVFTNSL
VCMGKIGEMDNNAKRRDARENKIIVMTLITSTVATRAGCDEALDVLAQMEQIGGV"
CDS
gene
repeat_region
Query Match 63.1%; Score 1341.4; DB 8; Length 122497;
Best Local Similarity 84.6%; Pred. No. 1.7e-132;
Matches 1640; Conservative 0; Mismatches 251; Indels 48; Gaps 10;
QY 2 tagaagagcgagtgatgcctcgcatcatgaagcggaggtacagagccggagggagcg 61
DB 69057 TAGGTAGTTTGGGGAGGCGAGATCATGAACGGCAGTACCAAAAGCCGGGAGCA 69116
QY 62 gcgagcgcgatgagcg-----atgggcctgcgcgagacaagatgagtgcgag 115
DB 69117 GCGGCGCGCGAGACAGCGCCGATATGGGCTGTCGAAGACAAAGGTGATGC---CGGGG 69173
QY 116 cggcggggaggggagggatggaagcagctgcgagcgcgctcgggtacaaggtgcgcg 175
DB 69174 CGCGCGGCGAGAGAGAGCAGCAGTCAAGAGCTCTGGCGCGCTCGGATCAAGTGCCTG 69233
QY 176 cctccgacatcgcgagcgtgagcagagcgtgagcagcgtgagagatgagcagggagatg 235
DB 69234 CCTCCGACATGGCCCACTGCGCGCAAGACTGACACAGTGGAGATGGCATGGGATGG 69293
QY 236 gcgagcgtgagcgcgcgcgcgcccccgacgaagcctgcgcacacacctgcgcagagca 295
DB 69294 GCGGCGTGAAGCGCCCGCGCGCGCGGATGACGGGTTGTCGCGACCTGCGCACAGGACA 69353
QY 236 ccgtgcaactacaacccacacgaactcgtgctcgtggtcgagagcatgctgcgaagctca 355
DB 69354 CCGTGCACATCAACCCCTCGACCTCTCTCTCGGTGAGAGCATGCTTCCGAGCTCA 69413
QY 356 acgagcgcgcgcgccccctcccgcccgcccgccgag-----ctaaagcctcacctct 409
DB 69414 ACGCGCGCTGCCCTTATCCCGCAGCGCGCGCGCTGCCGCAATGCTTCACCTCGT 69473
QY 410 ccacgcgacagggcagcg-----gcggctactcgcattcccgccctcgcgcactcct 463
DB 69474 CCACGTGTCACCGCGCGCGGCGGTGTAGCGGCTTCTTGAACCTCCACCGCTGCCGACTGT 69533
QY 464 ccagcagacatcagcgtcgtgcgacgcatccctcccgagcgcgcgacagcgcgcgcgcg 523
DB 69534 CGAGTACGACCTACCGCCCTCAAGCGCATCTCTTACCGGTGTGCGACAGG---CTGACC 69590
QY 524 acctgcgcgacatcgtgcgagatcccaagcgaatgcgaactgcgaggaacaacact 583
DB 69591 CGTGCGCTGTACTCGCGGAGGAGGACCAACGGATGCGCATGGCGGCGGACACACT 69650
QY 584 cgtcgtacatccctcccgatcgtcctc---cggtgggggagcgccgagagcctgtggtg 640
DB 69651 CGTGCTCTCATCGTGTCTTCTCTCTGCGGAGTGGGCGGCTTGTGTGTG 69710
QY 641 aggcgtcccgcccgcg-----tcgcgcgcgcgccaacgacgacgacgacgacgacgacg 694
DB 69711 AGGCTGCTCCGCGCGGACGCAAGAGGCGCGCGCGGCGCAATGCCCGCGCGGTTG 69770
QY 655 tcgtgtcgacaacgagagcgagatcggctgtggtgacagcgtcgtggtgcgagcg 754
DB 69771 TGGTGTGTACACGACGAGAGGCTGGGATCCGCTGTGTACACGCTGTGCGGTGCGCGG 69830
```



[illegible]

QY	1835	tgagctcggggtggaacaacgcgcgcctgtatgcgcacccctcgagcagggcgccggc-cgggc	1894
Db	70896	TCAGCTGGGCTGGCAACAGCGGCCGCTCTATCCGACACTGGGCGATGGCCGCTGCGCGAG	70955
QY	1895	cgtagctcgcgagttctg	1913
Db	70956	CGTGATCGCAAGTTTGG	70974
RESULT	8		
AX005806			
LOCUS	AX005806	2255 bp	DNA
DEFINITION	Sequence 15 from Patent WO9909174.		PAT
ACCESSION	AX005806		24-AUG-2000
VERSION	AX005806.1	GI:9928803	
KEYWORDS	.		
SOURCE	Zea mays.		
ORGANISM	Zea mays.		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta		
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC		
	clade; Panicoideae; Andropogoneae; Zea.		
	1 (bases 1 to 2255)		
	Harberd, N.P. and Peng, J.		
REFERENCE	Genetic control of plant growth and development		
AUTHORS	Patent: WO 9909174-A 15-25-FEB-1999;		
JOURNAL	HARBERD NICHOLAS PAUL (GB); PENG JINONG (GB)		
FEATURES	location/Qualifiers		
source	1..2255		
	/organism="Zea mays"		
	/db_xref="taxon:4577"		
BASE COUNT	334 a	817 c	737 g
ORIGIN			367 t

[illegible]











[illegible][illegible]

[illegible][illegible]

/organism="Zea mays subsp. mays"  
/cultivar="Iowa 73 (B73)"  
/isolate="B73"  
/sub\_species="mays"  
/db\_xref="taxon:4578"  
/chromosome="1"  
/product="gibberellin response modulator"  
<1..>800  
/note="d8"  
/codon\_start=3  
/product="gibberellin response modulator"  
/protein\_id="AK59925.1"  
/db\_xref="GI:14318163"  
/translation="PSADPSYDSAREPKRMRTGGSGTSSSSSSSSMDGRTSSVYE  
AAPATQSAANPAVPVYVDQFAGIRLVHLLACAEVAQDQENFSAALYKOIP  
MLASSQGAARKVATFGEALARRVYRFRPPDSSLDAFADLLHAFYESCPLYKF  
AHFTANQATLEAFNGRHHVDFGIRKQGMQWPAFLQALALRPGPPSFRLTGVGPPQ  
PDETDAIQVGMKLAQFAHTRVDFOYRGLVAATLADLEPMLQPED"

BASE COUNT 104 a 306 c 263 g 127 t  
ORIGIN

Query Match 29.0%, Score 617, DB 8; Length 800;  
Best Local Similarity 87.6%, Pred. No. 6.4e-56;  
Matches 700; Conservative 0; Mismatches 90; Indels 9; Gaps 2;

OY 512 cggcgccggccggaactgtccgcgcgactccgtgctggatcccaagagatgcgactgagc 571  
DB 1 cgcctgcggccgacccgctccacgacgactgcggcgccgaccccgacgagcagagcactggcg 60  
OY 572 ggaagcagcaccctcgtcgtcattccctcctcgtcgtcctcgttggg--gcgcagga 628  
DB 61 gggcgagcagcagctgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 120  
OY 629 gctctgtgtgtgagagctcgc 682  
DB 121 gctccggtgctcgaacgctgc 180  
OY 683 cgcctcgcgtcgtcgttgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 742  
DB 181 cggctccgctggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 240  
OY 743 tggcgtgc 802  
DB 241 tggcgtgc 300  
OY 803 agataccttgcgtgc 862  
DB 301 agatcccatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 360  
OY 863 ggcgaagccctgc 922  
DB 361 ggcgaagccctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 420  
OY 923 acgcgcgcctcgc 982  
DB 421 acgcgcgccttgc 480  
OY 983 tcgcgcacttcacgc 1042  
DB 481 tcgcgcacttcacgc 540  
OY 1043 acgtgcgtgacttcgacatcaagcagggagatgagtcgcgcgcgcgcgcgcgcgcgcgcgcgc 1102  
DB 541 acgtgcgtgacttcgacatcaagcagggagatgagtcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 600  
OY 1103 ccctcgc 1162  
DB 601 ccctcgc 660  
OY 1163 acgagaccgacgc 1222

DB 661 acgagaccgacgc 720  
OY 1223 gcgtgcacttcacgacgc 1282  
DB 721 gcgtgcacttcacgacgc 780  
OY 1283 tgctgcagccgc 1301  
DB 781 tgctgcacacgc 799

Search completed: January 23, 2002, 02:36:49  
Job time: 12789 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 23, 2002, 00:58:16 ; Search time 197.56 Seconds  
(without alignments)  
9221.584 Million cell updates/sec

Title: US-09-485-529-14

Perfect score: 2125

Sequence: 1 atagagaagcagtagctc.....tgatgagcagcaactccg 2125

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N.Geneseq.1101.\*  
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2125	100.0	2125	20	AA36279	Wheat Rht clone 5a
2	1956.2	92.1	2709	20	AA36275	Composite DNA sequ
3	1502.4	70.7	1768	20	AA36278	Wheat Rht clone C1
4	1379.8	64.9	1746	20	AA36285	Consensus cDNA seq
5	1317	62.0	2255	20	AA36280	Maize 1a1 genomic
6	420.4	19.8	770	20	AA36277	Rice EST D39460 se
7	382	18.0	1602	21	AA36274	Arabidopsis thaliaa
8	381.8	18.0	1964	18	AA36273	Arabidopsis thaliaa
9	380.2	17.9	1951	22	AAD06646	A. thaliana transc
10	380.2	17.9	1951	22	AAD05791	Arabidopsis thaliaa
11	380	17.9	1779	22	AAF25480	Nucleotide sequenc

12	378.4	17.8	1779	22	AA36281	Nucleotide sequenc
13	367	17.3	416	20	AA36283	Partial sequence o
14	359	16.9	436	20	AA36269	DNA sequence obtai
15	351.8	16.6	1643	18	AA36293	Arabidopsis thalia
16	347	16.3	1642	18	AA36294	Arabidopsis thalia
17	343	16.1	1642	18	AA36293	Arabidopsis thalia
18	329.4	15.5	1636	18	AA36290	Arabidopsis thalia
19	329.2	15.5	725	20	AA36276	Partial cDNA seque
20	326.6	15.4	1764	22	AA36261	A. thaliana transc
21	326.6	15.4	1764	22	AA36276	Arabidopsis thalia
22	322	15.2	377	20	AA36268	DNA sequence obtai
23	316.6	14.9	369	20	AA36258	DNA sequence obtai
24	312.2	14.7	453	20	AA36263	DNA sequence obtai
25	308	14.5	332	20	AA36255	DNA sequence obtai
26	297	14.0	324	20	AA36273	DNA sequence obtai
27	277.6	13.1	357	20	AA36259	DNA sequence obtai
28	267	12.6	425	20	AA36270	DNA sequence obtai
29	264.2	12.4	377	20	AA36256	DNA sequence obtai
30	249.2	11.7	309	20	AA36261	DNA sequence obtai
31	242	11.4	285	20	AA36271	DNA sequence obtai
32	238.4	11.2	511	20	AA36260	DNA sequence obtai
33	180	8.5	497	21	AA36267	Pinus radiata tran
34	178.4	8.4	230	20	AA36262	DNA sequence obtai
35	174.8	8.2	2151	18	AA36257	Maize ZCARECROW ZC
36	174.8	8.2	2151	21	AA36298	Maize ZCR gene par
37	171.8	8.1	3510	21	AA36299	Maize Scarecrow nu
38	171.6	8.1	399	20	AA36262	DNA sequence obtai
39	169.2	8.0	259	20	AA36272	DNA sequence obtai
40	165.8	7.8	302	20	AA36281	Partial sequence o
41	161	7.6	211	20	AA36257	DNA sequence obtai
42	155	7.3	969	21	AA36327	Maize GBRBT74 par
43	153.8	7.2	341	21	AA36316	Eucalyptus grandis
44	153.8	7.2	341	21	AA36358	Eucalyptus grandis
45	148	7.0	371	20	AA36282	Partial sequence o

#### ALIGNMENTS

RESULT 1  
ID AAX36279 standard; DNA; 2125 BP.  
XX AAX36279;  
AC AAX36279;  
XX  
DT 16-JUL-1999 (first entry)  
XX  
DE Wheat Rht clone 5a1 genomic sequence.  
XX  
KW Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;  
KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;  
KW paclobutrazol; ss.  
XX  
OS Triticum aestivum.  
XX  
PN WO9909174-A1.  
XX  
PD 25-FEB-1999.  
XX  
PF 07-AUG-1998; 98WO-GB02383.  
XX  
PR 13-AUG-1997; 97GB-0017192.  
XX  
PA (PLAN-) PLANT BIOSCIENCE LTD.  
XX  
PI Harberd NP, Peng J, Richards DE;  
XX WPI: 1999-181040/15.  
XX P-PSDB; AA102540.  
XX  
PT New Triticum aestivum polynucleotides - encode a polypeptide which provides inhibition of the growth of plants, which inhibition is antagonised by gibberellin, used to confer a dwarf phenotype

XX Disclosure: Fig 8a, 88pp; English.  
PS The specification describes polypeptides encoded by the Rht gene (and  
CC its homologues) that, when expressed in Triticum aestivum, inhibit  
CC growth of the plant. This growth inhibition is antagonised by  
CC gibberellin. The products can be used to provide Rht expression in  
CC plants, conferring a dwarf phenotype on a plant which is correctable  
CC by treatment with gibberellin. In addition, the products can be  
CC used to produce Rht mutant plants which are dwarfed compared with  
CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants  
CC may be made by knocking out Rht or the relevant homologous gene in  
CC the plant of interest. Plants may be made which are resistant to  
CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,  
CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds  
CC dwarf but let crop plants grow tall. The present sequence represents  
CC the wheat Rht clone 5a1 genomic sequence.  
XX  
S0 Sequence 2125 BP: 325 A; 768 C; 723 G; 309 T; 0 other;

Query Match 100.0%; Score 2125; DB 20; Length 2125;  
Best Local Similarity 100.0%; Pred. No. 1,1e-282;  
Matches 2125; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atagagagcgaggtagctcgagatcagatcaagcggaggtaccagagcgccggagagc 60  
DB 1 atagagagcgaggtagctcgagatcagatcaagcggaggtaccagagcgccggagagc 60  
OY 61 ggcgagcgagctgagcagatggctcgtccagaggaatgagtgtccggcgagcg 120  
DB 61 ggcgagcgagctgagcagatggctcgtccagaggaatgagtgtgtccggcgagcg 120  
OY 121 ggggggggggagagtgagcagagctgtcgagcgctcgaggtataaagtgtagccctcc 180  
DB 121 ggggggggggagagtgagcagagctgtcgagcgctcgaggtataaagtgtagccctcc 180  
OY 181 gacatgagcgagcgttgagcagaagatctggagcagctcgagatgagcattggagcgcg 240  
DB 181 gacatgagcgagcgttgagcagaagatctggagcagctcgagatgagcattggagcgcg 240  
OY 241 gtaggagcgagcgttgagcagaagatctggagcagctcgagatgagcattggagcgcg 300  
DB 241 gtaggagcgagcgttgagcagaagatctggagcagctcgagatgagcattggagcgcg 300  
OY 301 gactcaaaccccaacgagactgtcgtcttgagtgagagcattgtctcgagagctcaacgcg 360  
DB 301 gactcaaaccccaacgagactgtcgtcttgagtgagagcattgtctcgagagctcaacgcg 360  
OY 361 ccgagcgagccctccgagcccgcccgagctcaacgagctccacactctccacgctcaag 420  
DB 361 ccgagcgagccctccgagcccgcccgagctcaacgagctccacactctccacgctcaag 420  
OY 421 ggcagagcgagctactatctgactccgctccgtcgaactctctccagagcattcaagcg 480  
DB 421 ggcagagcgagctactatctgactccgctccgtcgaactctctccagagcattcaagcg 480  
OY 481 ctgagcgagatccctcccgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 540  
DB 481 ctgagcgagatccctcccgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 540  
OY 541 gtagcgagatcccaagcagatgagcactgtagcgagagcagcactctgtctcatcctcc 600  
DB 541 gtagcgagatcccaagcagatgagcactgtagcgagagcagcactctgtctcatcctcc 600  
OY 601 tctgtctctctctgtgtaggagcgagagctctgtgtgtagagctgagcgagcgagcg 660  
DB 601 tctgtctctctctgtgtaggagcgagagctctgtgtgtagagctgagcgagcgagcg 660  
OY 661 gccgagcgagcagagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 720  
DB 661 gccgagcgagcagagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 720

OY 721 atcgagctgtagcagcgagctgtagctgtagcgagagccggtgagcagagagactctcc 780  
DB 721 atcgagctgtagcagcgagctgtagctgtagcgagagccggtgagcagagagactctcc 780  
OY 781 gccgagagcgagctgtagagagatatacctgtgagcgagctcccaaggcgagcgagtg 840  
DB 781 gccgagagcgagctgtagagagatatacctgtgagcgagctcccaaggcgagcgagtg 840  
OY 841 cgaagagctcgagctactctcgagagagccctcgagcgagcgagctcttcgagctccg 900  
DB 841 cgaagagctcgagctactctcgagagagccctcgagcgagcgagctcttcgagctccg 900  
OY 901 cagcgagagcagctccctctcgagcgagccttcgagagctcttcagcgagctcttac 960  
DB 901 cagcgagagcagctccctctcgagcgagccttcgagagctcttcagcgagctcttac 960  
OY 961 gagtctgagctccctactcaagtgtgagcagcttcaacgagcagcagcagcttgagcg 1020  
DB 961 gagtctgagctccctactcaagtgtgagcagcttcaacgagcagcagcagcttgagcg 1020  
OY 1021 ttcgagcgagctgagcgagctgtagcagctgtagcagcttcaacgagcagcagctgag 1080  
DB 1021 ttcgagcgagctgagcgagctgtagcagctgtagcagcttcaacgagcagcagctgag 1080  
OY 1081 ccgagactctccagagccctcgagccttcgagcgagccttcgagctccgagcagc 1140  
DB 1081 ccgagactctccagagcccttcgagccttcgagcgagccttcgagctccgagcagc 1140  
OY 1141 ggcgtcgagcccccagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1200  
DB 1141 ggcgtcgagcccccagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1200  
OY 1201 gccagagctcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1260  
DB 1201 gccagagctcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1260  
OY 1261 ctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1320  
DB 1261 ctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1320  
OY 1321 cccgagagtaatcgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1380  
DB 1321 cccgagagtaatcgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1380  
OY 1381 gccctgagagagctctgagcagcagcagcagcagcagcagcagcagcagcagcagc 1440  
DB 1381 gccctgagagagctctgagcagcagcagcagcagcagcagcagcagcagcagcagc 1440  
OY 1441 gagcagagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1500  
DB 1441 gagcagagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1500  
OY 1501 tactatccacagcttgcagctccctcgagagcgagcagcagcagcagcagcagcagc 1560  
DB 1501 tactatccacagcttgcagctccctcgagagcgagcagcagcagcagcagcagcagc 1560  
OY 1561 gtcctatcgagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1620  
DB 1561 gtcctatcgagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1620  
OY 1621 gtagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1680  
DB 1621 gtagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1680  
OY 1681 cgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1740  
DB 1681 cgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1740  
OY 1741 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1800  
DB 1741 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1800  
OY 1801 gacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1860

[illegible]

## RESULT 2

ID AAX36275 standard; DNA; 2709 BP

AC AAX36275;

DT 16-JUL-1999 (first entry)

Composite DNA sequence of wheat Rht gene

KW Rht gene; homologue; *Triticum aestivum*; wheat; growth inhibition;  
KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;  
KW paclobutrazol; ss.

*Trilium aestivum.*

PN W09909174-A1

PD 25-FEB-1999

PF 07-AUG-1998; 98WO-GB02383.

PR 13-AUG-1997; 97GB-0017192.

PA (PLAN-) PLANT BIOSCIENCE LTD.

PI Harberd NP, Peng J, Richards DE;

DR WPI; 1999-181040/15.

PT New Triticum Aestivum polynucleotides - encode a polypeptide which

PT antagonist gibberellin, used to confer a dwarf phenotype

PS Disclosure; Fid 3a; 88pp; English.

The specification describes polypeptides encoded by the Rht gene (and its homologues) that, when expressed in *Triticum aestivum*, inhibit growth of the plant. This growth inhibition is antagonized by gibberellin. The products can be used to provide Rht expression in plants, conferring a dwarf phenotype on a plant which is correctable by treatment with gibberellin. In addition, the products can be used to produce Rht mutant plants which are dwarfed compared with wild-type, the dwarfing being gibberellin-insensitive. Taller plants may be made by knocking out Rht or the relevant homologous gene in the plant of interest. Plants may be made which are resistant to compounds which inhibit gibberellin biosynthesis, such as paclobutrazol, e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds dwarf but let crop plants grow tall. The present sequence represents

CC the composite DNA sequence of wheat Rht gene.  
XX  
SQ Sequence 2709 BP; 473 A; 900 C; 823 G; 482 T; 31 other;

Query Match	92.1%;	Score 1956.2;	DB 20;	Length 2709;
Best Local Similarity	96.8%;	Pred. No. 1.3e-259;		
Matches 2061; Conservative	0;	Mismatches 53;	Indels 15;	Gaps 7;

[illegible]

```

QY 961 gagctcggccctactcaagtctcgacactcaaccgcaacacagcactctctgagcgc 1020
    |||
Db 1085 gagctcggccctactcaagtctcgacactcaaccgcaacacagcactctctgagcgc 1144
QY 1021 ttcgcgcgctgcgcgcgcgtgcagctctcgacttcgcactcaagcagggatgcagtcg 1080
    |||
Db 1145 ttcgcgcgctgcgcgcgcgtgcagctctcgacttcgcactcaagcagggatgcagtcg 1204
QY 1081 cccgcacactctccagacccctcgccctcccgccgagccctccctccctccgctccacc 1140
    |||
Db 1205 cccgcacactctccagacccctcgccctcccgccgagccctccctccctccctccacc 1264
QY 1141 ggcgttcgccccccgcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1200
    |||
Db 1265 ggcgttcgccccccgcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1324
QY 1201 gcccaagctcgcgcacacacacacacacacacacacacacacacacacacacacacac 1260
    |||
Db 1325 gcccaagctcgcgcacacacacacacacacacacacacacacacacacacacacacac 1384
QY 1261 ctcgcgcgacactgagacgcttcacgtctgcacgacgagcagcagcagcagcagcagc 1320
    |||
Db 1385 ctcgcgcgacactgagacgcttcacgtctgcacgacgagcagcagcagcagcagcagc 1444
QY 1321 cccgcaggtacacgcgcgtacactacgtcttcagagatgcacacgcgcgtctgcgcagc 1380
    |||
Db 1445 cccgcaggtacacgcgcgtacactacgtcttcagagatgcacacgcgcgtctgcgcagc 1504
QY 1381 gcccttggaagaagc-tccctgggcacgcgtgcg-cgcgcgtgcgcgcgcgcgcgcgcgc 1437
    |||
Db 1505 gcccttggaagaagcttcttgggacccgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1564
QY 1438 gtcgg-aggcaggaagcgcagcagcagcagcagcagcagcagcagcagcagcagcagc 1496
    |||
Db 1565 gtcggaaacgcaggaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1624
QY 1497 gcaactactacacacacacacacacacacacacacacacacacacacacacacacac 1556
    |||
Db 1625 gcaactactacacacacacacacacacacacacacacacacacacacacacacacac 1684
QY 1557 cgaagctcactcggggcggtgcgtgcgtctctcgcgcgcgcgcgcgcgcgcgcgcgcgc 1616
    |||
Db 1685 cgaagctcactcggggcggtgcgtgcgtctctcgcgcgcgcgcgcgcgcgcgcgcgcgc 1744
QY 1617 cgaagctcactcggggcggtgcgtgcgtctctcgcgcgcgcgcgcgcgcgcgcgcgcgc 1676
    |||
Db 1745 cgaagctcactcggggcggtgcgtgcgtctctcgcgcgcgcgcgcgcgcgcgcgcgcgc 1804
QY 1677 agagcgacacagagacgcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1736
    |||
Db 1805 agagcgacacagagacgcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1864
QY 1737 cgcgcaccccggttcacacacacacacacacacacacacacacacacacacacacacac 1796
    |||
Db 1865 cgcgcaccccggttcacacacacacacacacacacacacacacacacacacacacacac 1924
QY 1797 cgcgcaccccggttcacacacacacacacacacacacacacacacacacacacacacac 1856
    |||
Db 1925 cgcgcaccccggttcacacacacacacacacacacacacacacacacacacacacacac 1983
QY 1857 cccgcgtatgcacacacacacacacacacacacacacacacacacacacacacacac 1916
    |||
Db 1984 cccgcgtatgcacacacacacacacacacacacacacacacacacacacacacacac 2043
QY 1917 gctgtaagtaacacacacacacacacacacacacacacacacacacacacacacacac 1976
    |||
Db 2044 gctgtaagtaacacacacacacacacacacacacacacacacacacacacacacacac 2103
QY 1977 cgcgcgaacgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2036
    |||
Db 2104 cgcgcgaacgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2163
QY 2037 cgcgtgaattgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2096

```

```

Db 2164 cgcgtgaattgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2223
    |||
QY 2097 gtcgaagagtgatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2125
    |||
Db 2224 gtcgaagagtgatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2252
    |||

RESULT 3
AAK36278
ID AAK36278 standard: cDNA; 1768 BP.
XX
AC AAK36278;
XX
XX
DT 16-JUL-1999 (first entry)
XX
DE Wheat Rht clone C15-1 cDNA sequence.
XX
KW Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;
KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;
KW paclobutrazol; ss.
XX
OS Triticum aestivum.
XX
PN WO9909174-A1.
XX
PD 25-FEB-1999.
XX
PF 07-AUG-1998; 98WO-GB02383.
XX
PR 13-AUG-1997; 97GB-0017192.
XX
PA (PLAN-) PLANT BIOSCIENCE LTD.
PI Harberd NP, Peng J, Richards DE;
XX
XX
DR WPI: 1999-181040/15.
DR P-PSDB: AAY02539.
XX
PT New Triticum aestivum polynucleotides - encode a polypeptide which
PT provides inhibition of the growth of plants, which inhibition is
PT antagonised by gibberellin, used to confer a dwarf phenotype
XX
PS Disclosure; Fig 7a; 88bp; English.
XX
CC The specification describes polypeptides encoded by the Rht gene (and
CC its homologues) that, when expressed in Triticum aestivum, inhibit
CC growth of the plant. This growth inhibition is antagonised by
CC gibberellin. The products can be used to provide Rht expression in
CC plants, conferring a dwarf phenotype on a plant which is correctable
CC by treatment with gibberellin. In addition, the products can be
CC used to produce Rht mutant plants which are dwarfed compared with
CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants
CC may be made by knocking out Rht or the relevant homologous gene in
CC the plant of interest. Plants may be made which are resistant to
CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,
CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds
CC dwarf but let crop plants grow tall. The present sequence represents
CC the wheat Rht clone C15-1 cDNA sequence.
XX
SQ Sequence 1768 BP; 368 A; 595 C; 527 G; 278 T; 0 other;

Query Match 70.7%; Score 1502.4; DB 20; Length 1768;
Best Local Similarity 99.9%; Pred. No. 1.3e-197;
Matches: 1503; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 622 gccagagctctgtgtgtgagcgtgccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 681
    |||
Db 1 gccagagctctgtgtgtgagcgtgccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 60
    |||
QY 682 ggcgtgcgcgtcgtcgtgtgtgcacgcagcagcagcagcagcagcagcagcagcagcagcagc 741
    |||

```

Db 61 ggcgtgcgtctgcgttggttcgacacgcagagccggagcttcgcgtgtgacgcgcgtg 120  
QY 742 ctggcgctggcgagagccgttgacagcagagaaaccttcgcgcgcggagagcgctggtagag 801  
Db 121 ctggcgctggcgagagccgttgacagcagagaaaccttcgcgcgcggagagcgctggtagag 180  
QY 802 cagatccctctgtgcgcgcgttcaccagagcgcgccgcatgcgaagctgcgcgcctcttc 861  
Db 181 cagatccctctgtgcgcgcgttcaccagagcgcgccgcatgcgaagctgcgcgcctcttc 240  
QY 862 ggcgagccctgcgcgcgcgttcctccgcttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 921  
Db 241 ggcgagccctgcgcgcgcgttcctccgcttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 300  
QY 922 ggcgagccctgcgcgcgcgttcctccgcttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 981  
Db 301 ggcgagccctgcgcgcgcgttcctccgcttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 360  
QY 982 ttcgcgcacttcacgcgcacacagccatcccttgagggcgttcgcgcgcgcgcgcgcgcgcgc 1041  
Db 361 ttcgcgcacttcacgcgcacacagccatcccttgagggcgttcgcgcgcgcgcgcgcgcgcgc 420  
QY 1042 cagcttcgcacttcgcgcacacagccatcccttgagggcgttcgcgcgcgcgcgcgcgcgcgc 1101  
Db 421 cagcttcgcacttcgcgcacacagccatcccttgagggcgttcgcgcgcgcgcgcgcgcgcgc 480  
QY 1102 ggccttcgccttcgc 1161  
Db 481 ggccttcgccttcgc 540  
QY 1162 gacgagacccgagcccttcgacagcgctggagcttgagagctgcgcgcgcgcgcgcgcgcgcgc 1221  
Db 541 gacgagacccgagcccttcgacagcgctggagcttgagagctgcgcgcgcgcgcgcgcgcgcgc 600  
QY 1222 cgcgttcgacttcacgataccgc 1281  
Db 601 cgcgttcgacttcacgataccgc 660  
QY 1282 atgcttcagcgagagcgagagagagacccgagagagagccgaggttaatcgccgtcaaac 1341  
Db 661 atgcttcagcgagagcgagagagagacccgagagagagccgaggttaatcgccgtcaaac 720  
QY 1342 tcaagcttcgagagtgacagcgcgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1401  
Db 721 tcaagcttcgagagtgacagcgcgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 780  
QY 1402 accgttcgc 1461  
Db 781 accgttcgc 840  
QY 1462 tccgagacatccctgcgacgcgcgttcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1521  
Db 841 tccgagacatccctgcgacgcgcgttcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 900  
QY 1522 tcccttcgagagggcgagcgcttcgc 1581  
Db 901 tcccttcgagagggcgagcgcttcgc 960  
QY 1582 gctccttcgc 1641  
Db 961 gctccttcgc 1020  
QY 1642 tgcgaacgttgctgc 1701  
Db 1021 tgcgaacgttgctgc 1080  
QY 1702 tggcggaacccggttcggcaaacgc 1761  
Db 1081 tggcggaacccggttcggcaaacgc 1140  
QY 1762 aagcagagcgagacgcgttcgtgc 1821  
Db 1141 aagcagagcgagacgcgttcgtgc 1200

QY 1822 aaggaagcgttcgctgc 1881  
Db 1201 aaggaagcgttcgctgc 1260  
QY 1882 cgccttcgc 1941  
Db 1261 cgccttcgc 1320  
QY 1942 gaggaacaacagcccgcgcgccgc 2001  
Db 1321 gaggaacaacagcccgcgcgccgc 1380  
QY 2002 ttgaagaagaagaagcgttaatgcatgtagtagcgtggaattgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2061  
Db 1381 ttgaagaagaagaagcgttaatgcatgtagtagcgtggaattgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1440  
QY 2062 atcgatcgagctacgggtggttcgc 2121  
Db 1441 atcgatcgagctacgggtggttcgc 1500  
QY 2122 tccg 2125  
Db 1501 tccg 1504

RESULT 4  
AA36285  
ID AAX36285 standard; cDNA; 1746 BP.  
XX  
AC AAX36285;  
XX  
DT 16-JUL-1999 (first entry)  
XX  
DE Consensus cDNA sequence of wheat Rht clone C15-1.  
XX  
KW Rht gene: homologue; Triticum aestivum; wheat; growth inhibition;  
KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;  
KW paclobutrazol; ss.  
XX  
OS Triticum aestivum.  
XX  
PN WO9909174-A1.  
XX  
PD 25-FEB-1999.  
XX  
PF 07-AUG-1998; 98WO-GB02383.  
XX  
PR 13-AUG-1997; 97GB-0017192.  
XX  
PA (PLAN-) PLANT BIOSCIENCE LTD.  
XX  
PI Harberd NP, Peng J, Richards DE;  
XX  
DR WPI; 1999-181040/15.  
XX  
PT New Triticum aestivum polynucleotides - encode a polypeptide which  
PT provides inhibition of the growth of plants, which inhibition is  
PT antagonised by gibberellin, used to confer a dwarf phenotype  
XX  
PS Disclosure; Fig 2a; 88pp; English.  
XX  
CC The specification describes polypeptides encoded by the Rht gene (and  
CC its homologues) that, when expressed in Triticum aestivum, inhibit  
CC growth of the plant. This growth inhibition is antagonised by  
CC gibberellin. The products can be used to provide Rht expression in  
CC plants, conferring a dwarf phenotype on a plant which is correctable  
CC by treatment with gibberellin. In addition, the products can be  
CC used to produce Rht mutant plants which are dwarfed compared with  
CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants  
CC may be made by knocking out Rht or the relevant homologue gene in  
CC the plant of interest. Plants may be made which are resistant to  
CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol.







**RESULT 6**

**AAX36277 standard; DNA: 770 BP.**

**AAX36277:**

**16-JUN-1999 (first entry)**

Rice EST D39460 sequence, homologous to wheat Rht gene.

Rht gene; homologue; Triticum aestivum; wheat; growth inhibition; antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis; paclobutrazol; rice; expressed sequence tag; EST; ss.

Oryza sativa.

WO9909174-AI.

25-FEB-1999.

07-AUG-1998: 98MO-GB02383.

13-AUG-1997: 97GB-0017192.

(PLAN-) PLANT BIOSCIENCE LTD.

Harberd NP, Peng J, Richards DE;

WPI: 1999-181040/15.

P-PDSB: AAY02538.

New Triticum aestivum polynucleotides - encode a polypeptide which provides inhibition of the growth of plants, which inhibition is antagonised by gibberellin, used to confer a dwarf phenotype

Claim 13; Fig 6a; 88pp; English.

The specification describes polypeptides encoded by the Rht gene (and its homologues) that, when expressed in Triticum aestivum, inhibit growth of the plant. This growth inhibition is antagonised by gibberellin. The products can be used to provide Rht expression in plants, conferring a dwarf phenotype on a plant which is correctable by treatment with gibberellin. In addition, the products can be used to produce Rht mutant plants which are dwarfed compared with wild-type, the dwarfing being gibberellin-insensitive. Taller plants may be made by knocking out Rht or the relevant homologous gene in the plant of interest. Plants may be made which are resistant to compounds which inhibit gibberellin biosynthesis, such as paclobutrazol e.g., to allow use of a gibberellin biosynthesis inhibitor to keep weed dwarf but let crop plants grow tall. The present sequence represents rice expressed sequence tag (EST) AAD39460, which is homologous to the wheat Rht gene.

Sequence 770 BP; 104 A; 252 C; 294 G; 120 T; 0 other;

Query Match	19.8%	Score 420.4	DB 20;	Length 770;
Best Local Similarly	77.5%	Pred.No. 9.1e-50;		
Matches 595; Conservative	0;	Mismatches 146;	Indels 27;	Gaps 6;

[illegible][illegible]

RESULT	7
AAC45745	
ID	AAC45745 standard; DNA; 1602 BP

AAC45745;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 47609.

KW Hybridisation assay; genetic mapping; gene expression control;

KW metabolic pathway; promoter; termination sequence; ss.

05 Arabidopsis thaliana

PN EP1033405-A2

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439

PR 25-FEB-1999; 99US-0121825

PR 09-MAR-1999; 99US-0123548



PR 23-MAR-1999: 99US-0125788.  
PR 25-MAR-1999: 99US-0126264.  
PR 29-MAR-1999: 99US-0126785.  
PR 01-APR-1999: 99US-0127462.  
PR 06-APR-1999: 99US-0128234.  
PR 08-APR-1999: 99US-0128234.  
PR 16-APR-1999: 99US-0129845.  
PR 19-APR-1999: 99US-0130077.  
PR 21-APR-1999: 99US-0130449.  
PR 23-APR-1999: 99US-0130510.  
PR 23-APR-1999: 99US-0130891.  
PR 28-APR-1999: 99US-0131449.  
PR 30-APR-1999: 99US-0132048.  
PR 04-MAY-1999: 99US-01332407.  
PR 05-MAY-1999: 99US-0132484.  
PR 06-MAY-1999: 99US-0132485.  
PR 06-MAY-1999: 99US-0132486.  
PR 07-MAY-1999: 99US-0132487.  
PR 11-MAY-1999: 99US-0132863.  
PR 14-MAY-1999: 99US-0134256.  
PR 14-MAY-1999: 99US-0134218.  
PR 14-MAY-1999: 99US-0134219.  
PR 14-MAY-1999: 99US-0134221.  
PR 14-MAY-1999: 99US-0134221.  
PR 14-MAY-1999: 99US-0134370.  
PR 18-MAY-1999: 99US-0134768.  
PR 19-MAY-1999: 99US-0134941.  
PR 20-MAY-1999: 99US-0135124.  
PR 21-MAY-1999: 99US-0135353.  
PR 24-MAY-1999: 99US-0135629.  
PR 25-MAY-1999: 99US-0136021.  
PR 27-MAY-1999: 99US-0136392.  
PR 28-MAY-1999: 99US-0136782.  
PR 01-JUN-1999: 99US-0137222.  
PR 03-JUN-1999: 99US-0137528.  
PR 04-JUN-1999: 99US-0137502.  
PR 07-JUN-1999: 99US-0137724.  
PR 08-JUN-1999: 99US-0138094.  
PR 10-JUN-1999: 99US-0138540.  
PR 10-JUN-1999: 99US-0138647.  
PR 14-JUN-1999: 99US-0139119.  
PR 16-JUN-1999: 99US-0139452.  
PR 16-JUN-1999: 99US-0139453.  
PR 17-JUN-1999: 99US-0139492.  
PR 18-JUN-1999: 99US-0139454.  
PR 18-JUN-1999: 99US-0139455.  
PR 18-JUN-1999: 99US-0139456.  
PR 18-JUN-1999: 99US-0139457.  
PR 18-JUN-1999: 99US-0139458.  
PR 18-JUN-1999: 99US-0139459.  
PR 18-JUN-1999: 99US-0139460.  
PR 18-JUN-1999: 99US-0139461.  
PR 18-JUN-1999: 99US-0139462.  
PR 18-JUN-1999: 99US-0139463.  
PR 18-JUN-1999: 99US-0139750.  
PR 18-JUN-1999: 99US-0139763.  
PR 21-JUN-1999: 99US-0139817.  
PR 22-JUN-1999: 99US-0139899.  
PR 23-JUN-1999: 99US-0140353.  
PR 23-JUN-1999: 99US-0140354.  
PR 24-JUN-1999: 99US-0140695.  
PR 28-JUN-1999: 99US-0140823.  
PR 29-JUN-1999: 99US-0140991.  
PR 30-JUN-1999: 99US-0141287.  
PR 01-JUL-1999: 99US-0141842.  
PR 01-JUL-1999: 99US-0142154.  
PR 02-JUL-1999: 99US-0142055.  
PR 06-JUL-1999: 99US-0142390.  
PR 08-JUL-1999: 99US-0142803.  
PR 09-JUL-1999: 99US-0142820.  
PR 12-JUL-1999: 99US-0142877.  
PR 13-JUL-1999: 99US-0143542.  
PR 14-JUL-1999: 99US-0143624.  
PR 15-JUL-1999: 99US-0144005.  
  
PR 16-JUL-1999: 99US-0144085.  
PR 16-JUL-1999: 99US-0144086.  
PR 19-JUL-1999: 99US-0144325.  
PR 19-JUL-1999: 99US-0144331.  
PR 19-JUL-1999: 99US-0144332.  
PR 19-JUL-1999: 99US-0144333.  
PR 19-JUL-1999: 99US-0144334.  
PR 19-JUL-1999: 99US-0144335.  
PR 20-JUL-1999: 99US-0144352.  
PR 20-JUL-1999: 99US-0144632.  
PR 20-JUL-1999: 99US-0144684.  
PR 21-JUL-1999: 99US-0144814.  
PR 21-JUL-1999: 99US-0145086.  
PR 21-JUL-1999: 99US-0145088.  
PR 22-JUL-1999: 99US-0145088.  
PR 22-JUL-1999: 99US-0145087.  
PR 22-JUL-1999: 99US-0145087.  
PR 22-JUL-1999: 99US-0145089.  
PR 22-JUL-1999: 99US-0145192.  
PR 23-JUL-1999: 99US-0145145.  
PR 23-JUL-1999: 99US-0145218.  
PR 23-JUL-1999: 99US-0145224.  
PR 26-JUL-1999: 99US-0145276.  
PR 27-JUL-1999: 99US-0145913.  
PR 27-JUL-1999: 99US-0145918.  
PR 27-JUL-1999: 99US-0145919.  
PR 28-JUL-1999: 99US-0145951.  
PR 02-AUG-1999: 99US-0146386.  
PR 02-AUG-1999: 99US-0146388.  
PR 02-AUG-1999: 99US-0146389.  
PR 03-AUG-1999: 99US-0147038.  
PR 04-AUG-1999: 99US-0147204.  
PR 04-AUG-1999: 99US-0147302.  
PR 05-AUG-1999: 99US-0147192.  
PR 05-AUG-1999: 99US-0147260.  
PR 06-AUG-1999: 99US-0147303.  
PR 06-AUG-1999: 99US-0147416.  
PR 09-AUG-1999: 99US-0147493.  
PR 09-AUG-1999: 99US-0147935.  
PR 10-AUG-1999: 99US-0148171.  
PR 11-AUG-1999: 99US-0148319.  
PR 12-AUG-1999: 99US-0148341.  
PR 13-AUG-1999: 99US-0148365.  
PR 13-AUG-1999: 99US-0148684.  
PR 16-AUG-1999: 99US-0149368.  
PR 17-AUG-1999: 99US-0149175.  
PR 18-AUG-1999: 99US-0149426.  
PR 20-AUG-1999: 99US-0149722.  
PR 20-AUG-1999: 99US-0149723.  
PR 20-AUG-1999: 99US-0149929.  
PR 23-AUG-1999: 99US-0149902.  
PR 23-AUG-1999: 99US-0149930.  
PR 25-AUG-1999: 99US-0150566.  
PR 26-AUG-1999: 99US-0150884.  
PR 27-AUG-1999: 99US-0151065.  
PR 27-AUG-1999: 99US-0151066.  
PR 27-AUG-1999: 99US-0151080.  
PR 30-AUG-1999: 99US-0151303.  
PR 31-AUG-1999: 99US-0151338.  
PR 01-SEP-1999: 99US-0151930.  
PR 07-SEP-1999: 99US-0152363.  
PR 10-SEP-1999: 99US-0153070.  
PR 13-SEP-1999: 99US-0153758.  
PR 15-SEP-1999: 99US-0154018.  
PR 16-SEP-1999: 99US-0154039.  
PR 20-SEP-1999: 99US-0154779.  
PR 22-SEP-1999: 99US-0155139.  
PR 23-SEP-1999: 99US-0155486.  
PR 24-SEP-1999: 99US-0155659.  
PR 28-SEP-1999: 99US-0156458.  
PR 29-SEP-1999: 99US-0156596.  
PR 04-OCT-1999: 99US-0157117.  
PR 05-OCT-1999: 99US-0157753.  
PR 06-OCT-1999: 99US-0157865.

PR	07-OCT-1999;	99US-0156029.
PR	08-OCT-1999;	99US-0156232.
PR	12-OCT-1999;	99US-0156369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

[illegible]

Db	977	atttgcgcgagcgatccatcgaactctgagcttgagatgacagagagattctgtggttaacactttag	1036
QY	1265	cggaccttgaaacccgtctcatctgcatcgacgggaagcgagagaaccggaaacggagagcccg	1324
Db	1037	ctgatacttgatgccttcgatactgctgagc-----tttagcaaanagttagatg	1081
QY	1335	aggtatcgcgcgtcaactcagactcttcgagatgcaaccggtctgtctgcgcgaacccggcgcc	1384
Db	1082	aatctgtgtcggtttaactctgctttctcgagcttcacagaactctttggagacactgtgtgcga	1141
QY	1385	tggagaaagtctcttgagcaacggtgcgcgcgcgtgcgcgcgcagatctgcaacatcggtgtgcga	1444
Db	1142	tcgataaaggtctctgtgtgtgtgtgaatcagaattaaacggagagattctcaactgtgtgtgcgc	1201
QY	1445	aggaagcgcaatcacaaactccgcgcacatctcttgagacgcgtctcaacgagttctcgactact	1504
Db	1202	aggaatcgaacacaataaataatgcgaattctcttaagatcgtttacatcgtatgctgtgcatact	1261
QY	1505	actccacatcttcgaattccctcgcgagagcgcgacgctccgcgcgcgcgcacatccgaactct	1564
Db	1262	actcgacgtgtgttgactcgtgtgaaag-----	1289
QY	1565	catcggggcgtcgtcgtctcgtctcgtccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1624
Db	1290	-----tgtaaccgagtggtcaagacaaggtcatgctcgaggttt	1327
QY	1625	accctcgc	1684
Db	1328	acttgggtgaaacagatctcgaacacgttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1387
QY	1685	acgagacgtctgagtcagcttgccgc	1744
Db	1388	atgaaacgttgaagtcagcttgagagaaacgcgtctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1447
QY	1745	tgggctccaatgtctctacaagcagagcgagacgctgtctgtgcgcgcgcgcgcgcgcgcgcgcgc	1804
Db	1448	ttgtgtcgaatgcgttttaagacaagcagatgctcttggctctgtctcaacgcgcgcgcgcgcgcgc	1507
QY	1805	gctcaaaagtctgagagagaaagagcgtcgtctgagcgtctgtgagctgtgagacacgcgcgcgcgcgc	1864
Db	1508	gttaccggtgtgagagagatgacgcgctgtctctcaatgctgtgagctgtgagacacgcgcgcgcgc	1567
QY	1865	tcgcacacctgc	1920
Db	1568	tagccacctgc	1593
RESULT 8			
AAT91937			
ID	AAT91937	standard; DNA; 1964 BP.	
XX	XX		
AC	AAT91937;		
XX	XX		
DE	19-MAR-1998	(first entry)	
XX	XX		
XX	XX	Arabidopsis thaliana gibberellin insensitivity gene gai.	
KM	KM	Gibberellin insensitivity; gai; plant growth inhibition;	
KM	KM	dwarf phenotype; lodging resistance; increased yield;	
KM	KM	flowering regulation; bolting inhibition; spinach; lettuce;	
KM	KM	antibody; identification; probe; primer; antisense; sense;	
KM	KM	expression regulation; co-suppression; rice;	
KW	KW	Bakane disease resistance; ss.	
OS	OS	Arabidopsis thaliana.	
XX	XX		
PN	PN	MO9729123-A2.	
XX	XX		
PD	PD	14-AUG-1997.	
XX	XX		
PF	PF	12-FEB-1997; 97MO-GB00390.	
XX	XX		





ID	AAID05791	standard; cDNA; 1951 BP.
XX	AAID05791;	
XX	31-JUL-2001	(first entry)
XX	Arabidopsis thaliana	transcription factor, G308 cDNA.
XX	Transcription factor; biochemical characteristic; controlling element;	
XX	structural characteristic; developmental characteristic; gene therapy;	
XX	agricultural biotechnology; plant trait modification; ss.	
XX	Arabidopsis thaliana.	
XX	Key	Location/Qualifiers
XX	CDS	196..1794
XX	FT	/tag= a
XX	FT	/product= "transcription factor, G308"
XX	NC0200136597-A1.	
XX	14-NOV-2000;	2000MO-US31344.
XX	17-NOV-1999;	99US-0166228.
XX	17-APR-2000;	2000US-0197899.
XX	22-AUG-2000;	2000US-0227439.
XX	(MEND-) MENDEL BIOTECHNOLOGY INC.	
XX	(CREE/) CREELMAN R.	
XX	(YUGG/) YU G.	
XX	(ADAM/) ADAM L.	
XX	(RIEC/) RIECHMANN J L.	
XX	(HEAR/) HEARD J.	
XX	(SAMA/) SAMAHA R.	
XX	(PIIG/) PILGRIM M.	
XX	(PINE/) PINEDA O.	
XX	(JIAN/) JIANG C.	
XX	Creelman R, Yu G, Adam L, Riechmann JL, Heard J, Samaha R;	
XX	Pilgrim M, Pineda O, Jiang C;	
XX	WPI: 2001-335999/35.	
XX	P-PSDB; AAED1907.	
XX	Nucleic acids encoding plant transcription factor polypeptides, useful	
XX	for altering the biochemical characteristics of plants e.g. corn,	
XX	potato and cotton plants -	
XX	Claim 4; Page 112-113; 127pp; English.	

[illegible]

DB 1517 ttacttggtaaacagatctgaacgtttgtgtgtatgtgacgtgacccagttgagc 1576  
QY 1682 gccacgagcgctggcgacgttgcggaaccggttggcaaccgggttcgagaccgttc 1741  
DB 1577 gtaataaacgcttgaatcagttgaggaaccggttcggtctcgtggttgcggtccgac 1636  
QY 1742 acctgggcccacgtcctaagcaagcagacgacgctgtcgtcgtccttcgcgcgagcg 1801  
DB 1637 atattgttcgaaatgcttaagcaagcgagatgtcttgcgtcctcgttcaacgagcggtc 1696  
QY 1802 acggctcaacaagtgtgaggaagaagcgtcgtcgtcgtggtgtgcaacgcgcgcgc 1861  
DB 1697 aggtgtatcgggttgaggaagatgaacggtctcattgtgtgtgttgcaacacgcgcgc 1756  
QY 1862 tgatcgccacgtcgatggcgctggcc 1890  
DB 1757 tcatagcaaccctcgcttggaaactctcc 1785

## RESULT 11

AAF25480

ID AAF25480 standard: DNA; 1779 BP.

XX AAF25480;

XX 15-MAY-2001 (first entry)

XX Nucleotide sequence of a wildtype GRAS protein.

XX GRAS protein; BZH gene; transcription factor; gibberellin; morphogenesis;  
XX plant development; dwarf plant; crucifer; ss.

XX Brassica napus.

XX Key Location/Qualifiers

XX FT CDS 60..1778

XX FT /tag= a

XX PN /product= "GRAS"

XX MO200109356-A1.

XX 08-FEB-2001.

XX 02-AUG-2000; 2000MO-FR02216.

XX 02-AUG-1999; 99FR-0010023.

XX (INRG ) INRA INST NAT RECH AGRONOMIQUE.

XX Renard M, Deloume R, Barret P, Brunel D, Froger N, Tanguy X;

XX MPI; 2001-182964/18.

XX P-PSDB; AAB31883.

XX New mutant nucleic acid encoding modified GRAS family protein, used to  
XX produce dwarf transgenic plants

XX Example 1; Page 13-15; 28pp; French.

XX The present sequence encodes a wild type plant protein of the GRAS  
XX family. The specification describes a mutant allele of the BZH gene,  
XX which contains a G1695A mutation resulting in the mutation E546K in the  
XX protein. GRAS proteins are transcription factors implicated in  
XX regulation of the response to gibberellins and thus in control of  
XX morphogenesis and plant development. The mutant GRAS protein is  
XX used to produce dwarf plants, specifically crucifers. Dwarf plants may  
XX be sown earlier (increasing nitrate accumulation without risking  
XX excessive stem growth during winter), and have better resistance to  
XX cold and lodging. They are also easier to harvest and allow for better  
XX monitoring of the crop.

XX Sequence 1779 BP; 418 A; 430 C; 476 G; 455 T; 0 other;

Query Match 17.9%; Score 380; DB 22; Length 1779;  
Best Local Similarity 60.1%; Pred. No. 2.7e-44;  
Matches 731; Conservative 0; Mismatches 410; Indels 75; Gaps 3;

QY 678 gcccgcgctgcggtctgcgtgtgttcgacacgagagcgaggttcggtgtgcaacgc 737  
DB 629 gcaactcgtccttaactcgtgacgtcgcaggaacagagtgctgtagtccacga 688  
QY 738 gctgtcgtcgttcgagcgagcggttcagcagagaaaccttcgcgcgcgagagcgctg 797  
DB 689 gcttatggtcgtcgttgaagccgttcagagcagcaacttgatctcagagtggtctcgt 748  
QY 798 gaagcaagatacccttgcgtgcgcgcgtcccaaggcgcgagtcgcaaggttcgcgccta 857  
DB 749 taagcaagatgtgttcttcgtgcgtctctcctcaacgcgagacatgaggaagtcgcaagta 808  
QY 858 ctgcgagagcgccctgcgcgcgcgcttccttcgcgttcgcgcgcgcgcgcgcgcgcgc 917  
DB 809 ctgcgcgagagctcgc 862  
QY 918 cctgcagcgcgcttcgc 977  
DB 863 gatcgatacctcttaccgatacctcctcagatcgcacttcagagactgtcccttaacct 922  
QY 978 caagttcgccacttcaaccgcgaaccagagcactctcgtgagcggttcgcgcgcgcgcgc 1037  
DB 923 caagttcgccacttcaaccgcgaaccagagcactctcgtgagcggttcgcgcgcgcgcgc 982  
QY 1038 cgtgcaggtcgtcgcgacttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1097  
DB 983 agtccagatcagatcgttctcgtatgaccaaaggtcttcgtgcccgcgcgcgcgcgcgcgc 1042  
QY 1098 cctgc 1157  
DB 1043 ccttggttgagaggaag 1102  
QY 1158 gccgagcagagacgc 1217  
DB 1103 ggcgagatacctcgc 1162  
QY 1218 catcgcgtcgc 1277  
DB 1163 gatcagcgtcgc 1222  
QY 1278 gttcatgctgc 1337  
DB 1223 ctcgatacgtcgc 1267  
QY 1338 caactcagcttcgc 1397  
DB 1268 taactcgttctcgc 1327  
QY 1398 ggcgacgcgtgc 1457  
DB 1328 cgcgcgttcgc 1387  
QY 1458 caactcgc 1517  
DB 1388 taacgctcgc 1447  
QY 1518 cgaatccctcgc 1577  
DB 1448 tgattccttggaag----- 1462  
QY 1578 tgcctcctcgc 1637  
DB 1463 -----tgctcgc 1513  
QY 1638 gatcgcgaacggtgtgtgc 1697  
DB 1514 gattgcgaactcgtgtgtgc 1573

```

OY      1698 ccaaggcgggaaccgagcttggcgaaacgcgggttcgtagaacgccaactccgggtccataatgc 1757
        ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       1574 tcaatgtctcacgcgglttcgtttcttgcticgagtcltgcgccggcgcatctcgggtctaaccgc 1633
        ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      1758 ctacaagaacgagcagcacgcctgtctgcgcgtctlctgcgcgcggcgacagcgtacaagaattga 1817
        + ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       1634 gtttaagaacagcgagtacgcgtcttttgtaccttgtaaattgaaaggcgaagttaatcgtgtga 1693
        ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      1818 gagagaaaggaagctctgcctcgcacgcctggggtgacacacgcgccgcctgatccacactgcgc 1877
        ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       1694 ggagaataaatggtgtgttgatgtgttgatggtgcacacactgcacgcgtcataaccaactccgc 1753
        ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      1878 atggcgcctgycgccggg 1893
        ||| | | | | | | |
Db       1754 ttggaagctctcgcg 1769
        ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT_12
ID      AAF25481 standard; DNA; 1779 BP.
XX      AAF25481;
AC      AAF25481;
DF      15-MAY-2001 (first entry)
DE      Nucleotide sequence of a mutant GRAS protein.
XX      DE
KM      GRAS protein; BZH gene; transcription factor; gibberellin; morphogenesis; plant development; dwarf plant; crucifer; ss.
OS      Brassica napus.
FX      Key Location/Qualifiers
FT      CDS 60..1778 /tag= a
FT      FT /product= "GRAS"
FN      WO200109356-A1.
PN      XX
PD      08-FEB-2001.
PE      02-AUG-2000; 200OMO-FR02216.
PR      02-AUG-1999; 99FR-0010023.
PA      (INRG ) INRA INST NAT RECH AGRONOMIQUE.
PI      Renard M., Delourme R., Barret P., Brunel D., Froger N., Tanguy X.: WPTI; 2001-182964/18.
DR      P-PADB; AAB31884.
DT      New mutant nucleic acid encoding modified GRAS family protein, used to produce dwarf transgenic plants - PT
PS      Example 1; Page 18-20; 28pp; French.
PX      CC The present sequence encodes a mutant plant protein of the GRAS family. The mutant allele of the BZH gene contains a G1695A mutation resulting in the mutation E546K in the protein. GRAS proteins are co-transcription factors implicated in regulation of the response to gibberellins and thus in control of morphogenesis and plant development. CC The mutant GRAS protein is used to produce dwarf plants, specifically crucifers. Dwarf plants may be sown earlier (increasing nitrate accumulation without risking excessive stem growth during winter), and have better resistance to cold and lodging. They are also easier to harvest and allow for better monitoring of the crop. CX
SQ      Sequence 1779 BP; 419 A; 430 C; 475 G; 455 T; 0 other.;
Query Match          17.8%; Score 378.4; DB 22; Length 1779; Best Local Similarity   60.0%; Pred No. 4.4e-44;
```

[illegible]

OY	1758	ctacaagcagcgacgacgcgtccgtcccttcgccggcgcgcaagcgtacaagtgtga	1817
Db	1634	gtctaaagaacgagctacgacttttgaccttctttaatgatgaaggcgaagttaatcgtgtga	1693
OY	1818	ggaagaaggaagcgtcccccagcgtctgggtgacacacgcgcgcgtgatcgcacctcggc	1877
Db	1694	gaagataattggtgttttgatgttgtgagctgcacactcagcacgtcataaacactctcgc	1753
OY	1878	atggcgcctggcgagg	1893
Db	1754	ttagaagctctcgcgcg	1769
 RESULT 13 AAAX36283 ID AAAX36283 standard; DNA; 416 BP. XX XX AC AAX36283; XX DT 16-JUL-1999 (first entry) XX DE Partial sequence of the wheat rht-10 allele. XX RW Rht gene; homologue; Triticum aestivum; wheat; growth inhibition; KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis; XX paclobutrazol; ss. XX OS Triticum aestivum. XX PN W09909174-A1. XX PD 25-FEB-1999. XX PE 07-AUG-1998; 98W0-G802383. XX PR 13-AUG-1997; 97GB-0017192. XX PA (PLAN-) PLANT BIOSCIENCE LTD. XX PI Harberd NP, Peng J, Richards DE; XX PT WPl: 1999-181040/15. DR P-PSDB; MAY02544. XX XX New Triticum Aestivum polynucleotides - encode a polypeptide which PT provides inhibition of the growth of plants, which inhibition is PT antagonised by gibberellin, used to confer a dwarf phenotype XX XX Disclosure: Fig 12a: 88bp; English.			

Query Match	17.3%;	Score 367;	DB 20;	Length 416;
Best Local Similarity	91.4%;	Pred. No. 1.9e-42;		
Matches 416; Conservative	0;	Mismatches 0;	Indels 39;	Gaps 1.

QY	151	gcgcgccctcggggtacaaagctgcgcgcgtccctccgaacatctgagcaactgtgagcgacgaagactgagag	210
Db	1	gcgcgcctcgggtacaaagctgcgcgcgtccctccgaacatctgagcaactgtgagcgacgaagactgagag	60
QY	211	cagctcgaagatgagccatctgagatctgagcgcgctgagcgcccgagcgccgcgcccccgacgaacagc	270
Db	61	cagctcgaagatgagccatctgagatctgagcgcgctgagcgcccgagcgccgcgcccccgacgaacagc	120
QY	271	ctcgccaccacactccgcagcgacacgctgtcactataccccacgcgaactgtctcttg	330
Db	121	ctcgccaccacactccgcagcgacacgctgtcactataccccacgcgaactgtctcttg	180
QY	331	gtcgcgagcagctgcgtctcgagactcaacgcgcgcgccgcgccctccgcgcgcgcgcgcag	390
Db	181	gtcgcgagcagcagctgcgtcgag-----ag 201	
QY	391	ctcaacgctccacactccctccacgcgtcaacggcagcgcgactcatctgcattctccgcc	450
Db	202	ctcaacgctccacactccctccacgcgtcaacggcagcgcgactcatctgcattctccgcc	261
QY	451	tcgcgtcgaactctccagagatcatctaaagcgctgagcgccagatctccctcccgccggcgagcg	510
Db	262	tcgcgtcgaactctccagagatcatctaaagcgctgagcgccagatctccctcccgccggcgagcg	321
QY	511	acgcgcgcgcgcgcgaactctccgcgcgaactcgtctgagggatcccaagcggaatctgcacatgc	570
Db	322	acgcgcgcgcgcgcgaactcgtctgagggatcccaagcggaatctgcacacatgc	381
QY	571	gggagcagacactcgtctcatctctctctctgc 605	
Db	382	gggagcagacactcgtctcatctctctctctctgc 416	

[illegible]





```

Db 1010 gaattcgtccaccggaacccggaataattcgaatcattcatgaagtgtggtgtaagcttg 1069
QY 1202 ccagattcgccagaccacccatccgctcgaacttcgaatccgagccctctgcgcgcacgc 1261
Db 1070 cccatttaagctgaagcgatccagctcgttgagttcgatccagagattgtgtgaacactt 1129
QY 1262 tcgcggaacctgagccggttcacatgctgcagccggaaggcgaggaacccgaacgaagagc 1321
Db 1130 tagctgattcttgatgcttcgagcttgagc-----ttagaccagtgaga 1174
QY 1322 ccgagagtaatcgccgtccaactcgaatcttcgagatgcacccggtctgcgcagcccgcg 1381
Db 1175 ttgaaactgtgtcggttcaactcgtgttcgagcttcacaaagctcttggaacgaacctgtg 1234
QY 1382 cccctggagaaggctcctgggacacggtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1441
Db 1235 cgatcgaataaggcttcgtgtgtgtggaatcaagatcaaacggagatttcacatgtgtgtg 1294
QY 1442 agcagagagcggaatcaacaactccggaacatcccttgagccgcttcacccgaagctctgcact 1501
Db 1295 agcaggaatcgaacacataataatgccgaatttcttaagatcgtttactgagtcgttgcat 1354
QY 1502 actactcaacatggtcgaatccctcgagggcggaagctccgcgcgcgcgcacatccgaag 1561
Db 1355 attactcgacgttgtgtgactcgttggaag----- 1385
QY 1562 tctcatcgggggtcgtcgtgtcctcctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1621
Db 1386 -----tgtaacgaagtgtcaagaacaaggctcaatgtcgaag 1420
QY 1622 tgtacctcgccggaatcgtcaacggtgtgtgcctgcgaggggcgagcgacagaagc 1681
Db 1421 tttaactgtgtataacagaatctgcgaacgtgtgtgtgtgtgacttgacaggttgagc 1480
QY 1682 gccacgagacgctgtggcgaagtgtgcggaacccgctgtggcaacggttcgagaacctcc 1741
Db 1481 gtcatgaaacgttgagtcagtgaggaacccggttcggtcgtgtgtgtgtgtgtgtgtgtgt 1540
QY 1742 acctgggtcccaatgcttcaacagcgagcgacgctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1801
Db 1541 atatcgtgtcgaatcgcgttaagcaagcgagcatgcttctgtgtgtgtgtgtgtgtgtgtgt 1600
QY 1802 acggtacacaaggctgagaggaaggaagagcgtgcctgaagcgtggg 1844
Db 1601 aggttatcgggtgtgaggaagtgagcgcgtgtcatgtgtggg 1643

```

Search completed: January 23, 2002, 02:37:26  
 Job time: 5950 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 22, 2002, 23:01:56 ; Search time 1805.74 Seconds  
(without alignments)  
12645.663 Million cell updates/sec

Title: US-09-485-529-14  
Perfect score: 2125  
Sequence: 1 atagagagcagcagtagctc.....tgtagtgacgacgactccg 2125

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapect 1.0

Searched: 11351937 seqs, 5372889281 residues  
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1: em\_estfun:\*  
2: em\_estin:\*  
3: em\_estom:\*  
4: em\_estpl:\*  
5: em\_estba:\*  
6: em\_estro:\*  
7: em\_estov:\*  
8: em\_hc:\*  
9: em\_hc1:\*  
10: gb\_est1:\*  
11: gb\_est2:\*  
12: gb\_hc:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rtd:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	563.8	26.5	778	11	BG308975 HVSMC000
2	516.4	24.3	597	11	BF620181 HVSMC001
3	503	23.7	851	11	BF254247 HVSMC000
4	464.2	21.8	563	11	BF588097 FMI_37_F0
5	445	20.9	593	10	AI881894 606074608
6	421.2	19.8	551	10	AI770444 606052F10
7	382.2	18.0	443	11	BF655617 FMI_46_D1
8	381.4	17.9	559	11	BF618807 HVSMC000
9	365.2	17.2	440	11	BI098997 IPI_36_A1
10	335.2	15.8	400	10	BE414891 MWL001_H0
11	329	15.5	481	11	BI417150 949053608
12	309.2	14.6	844	11	BF268018 HV_CEA001

13	307.4	14.5	527	10	BE595338	BE595338 P11_48_G1
14	284.2	13.4	596	10	AU065169	AU065169 AU065169
15	273.4	12.9	524	10	BE329503	BE329503 so6b12.Y
16	263.8	12.4	563	11	BG043071	BG043071 st91h01.Y
17	258.8	12.2	543	11	BG155663	BG155663 sa64a06.Y
18	257.4	12.1	579	10	AW704479	AW704479 SK35e12.Y
19	256.2	12.1	467	10	AU222715	AU222715 AU222715
20	252	11.9	475	10	BE364489	BE364489 P11_14_F0
21	251.4	11.8	658	11	BF587541	BF587541 FMI_37_F0
22	244.2	11.5	588	10	AI665907	AI665907 606003E05
23	243	11.4	577	10	BE196286	BE196286 HVSMC009
24	242.2	11.4	981	11	BF201178	BF201178 WHE0986_A
25	224	10.5	268	10	BE454560	BE454560 HVSMC009
26	223.4	10.5	743	10	AW774515	AW774515 EST333666
27	220	10.4	450	10	AW396192	AW396192 sh02e12.Y
28	219.8	10.3	626	11	BF009011	BF009011 ss72c03.Y
29	217.6	10.2	592	11	BG13881	BG13881 EMI_60_F1
30	215.6	10.1	494	11	BG510374	BG510374 sac76h08.Y
31	211.8	10.0	408	10	AV423888	AV423888 AV423888
32	211.8	10.0	714	10	AW348855	AW348855 GM210010A
33	211.2	9.9	511	11	BG790472	BG790472 sae58a03.Y
34	209	9.8	416	11	BG051785	BG051785 FMI_58_DO
35	208.4	9.8	379	10	AV420689	AV420689 AV420689
36	207.2	9.8	452	10	AI416823	AI416823 sa18d06.Y
37	206.6	9.7	430	10	AV413998	AV413998 AV413998
38	201.4	9.5	700	11	BG587404	BG587404 EST489176
39	196.4	9.2	587	11	BE805895	BE805895 ss62g03.Y
40	190.2	9.0	473	10	AW734476	AW734476 SK94h05.Y
41	186.2	8.8	700	10	AL506960	AL506960 AL506960
42	182	8.6	626	10	BE204113	BE204113 EST396789
43	180.6	8.5	1064	10	BE035220	BE035220 MO01E08
44	177.2	8.3	420	10	BE610936	BE610936 sq68e11.Y
45	173.4	8.3	671	11	BG43698	BG43698 GA_Ea002

#### ALIGNMENTS

RESULT 1  
BG308975 778 bp mRNA EST 22-FEB-2001  
LOCUS HVSMC0001c18f Hordeum vulgare seedling shoot EST library  
DEFINITION HVCNMA0003 (Etiolated and unstressed) Hordeum vulgare cDNA clone  
HVSMC0001c18f, mRNA sequence.  
ACCESSION BG308975  
VERSION BG308975.1 GI:13109822  
KEYWORDS EST.  
SOURCE Hordeum vulgare  
ORGANISM Hordeum vulgare  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae  
; Trilliceae; Hordeum.  
REFERENCE Wing,R., Close,T.J., Kleinbofs,A., Wise,R., Begum,D., Frisch,D., Yu  
1 (bases 1 to 778)  
Y., Anderson,H., Dale,J., Henry,D., Kennodie,S., Palmer,M., Ranbo  
,T., Saski,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D. and  
Wood,T.  
Development of a genetically and physically anchored EST resource  
for barley genomics  
Unpublished (2000)  
CONTACT: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Seg primer: AATTAACCTCACTAAGG  
High quality sequence stop: 690.  
Location/Qualifiers  
1..778  
/organism="Hordeum vulgare"  
/cultivar="Morex"

```

/db_xref="taxon:4513"
/clone="HVSMEC0018M1f"
/clone_lib="Hordeum vulgare seedling shoot EST library
HVCNMA0003 (Etiolated and unstressed)"
/tissue_type="Seedling shoot"
/lab_host="TUC121"
/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI. For
more details on library preparation and sequence analysis
see http://www.genome.clemson.edu/projects/barley/ To
order a clone see http://www.genome.clemson.edu/orders"
BASE COUNT      96 a      295 c      261 g      124 t      2 others
ORIGIN

```

```

Query Match      26.5%: Score 563.8; DB 11; Length 778;
Best Local Similarity 87.2%; Pred. No. 2.4e-68;
Matches 695; Conservative 0; Mismatches 74; Indels 28; Gaps 6;

QY 154 ggcgtcgagtaacaagtgccgctccgacatgagcgtgagcagagaagctgagcag 213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 GCGGGGGGGTGGAGAGGGGCGGCGTCCGACATGGCGGACGTGGCGCAGAGCTGAGCAG 60

QY 214 ctgagatgagcctatgaggaatgaggcgctgaggcgccgagcccgccgagacagcttc 273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 CTCGAGATGGCCATGGGATGGG-----CGGCCCGCCCGCCGACGAGCGCTTC 108

QY 274 gcccaccaccctcgccagcagacgctgactacaaccaccagacgtgctgtctggtc 333
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109 GCGACCCACCTCCGACGAGCAGCAGCTCCACACACCCACGACCTCTCTGCTGGGGG 168

QY 334 gagagatgctgctgagagctaaagcgccgagccgcccctccgagccgccc---ccgag 390
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 169 GAGAGCATGCTGTCGAGACTCAACGCGCGCGCGCCCTCCCGCGCGCGCGCGCAG 228

QY 391 ctcaagcctcaactctctccacgctcagcgagcgagcgagctacttctccgccc 450
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 229 CTCACGCTCCACCTCTCCACGCTCAAGGGGCGGCGGATCTTGATCTCCGCGCC 288

QY 451 tccgtgactctctccagagacatctacgctgagcgagccgacccctcccgagcgagcg 510
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 289 TCTGTGCACTCTCCAGAGCAGCAGCAGCTACGCGCCCGATCATCTCCGCGCGCTC--- 345

QY 511 aagcgccgagcagcagctgctcgagcagctgagcagctcccaagcagagtgagcagctg 570
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 346 ---GCGCGCGCGACCTCTCCGCTGACTCGTCCGAGACCCCAACGAGATGCGGCACTGCG 402

QY 571 gggagcagacactcgctgctatctctcctcctcgtctctc---cggtgaggcgagcg 627
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 403 GGCAGCACACGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 462

QY 628 agctctgtgtgagagcgtgagccgagctgagcgagcagcagcagcgagcgagcgctg 687
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 463 AGCTCTGTGTGAGAGGCTGTCTCCGCGGTGGGCTGCGGC---TGTGCGCGCCCGCTG 519

QY 688 ccggtcgctgctgctgagcagcagagcgagcagctgagctgagcagcgctgagcg 747
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 520 CCGGTCTGTGTGAGAGCAGCAGCAGGCGGAGATTCGGCTGTGTCAGCGCTGTGCGC 579

QY 748 tgcgagcagcagcgtgagcagagaaactctccgagcgagcgagcgctggtgaagcagata 807
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 580 TGCAGGAGGCGCTGTGACAGAGAACTCTCGGCGCGGAGGCGCTGTGTAAGCAGATA 639

QY 808 ccccttgagcagcgtgagcagagcgagcgagctgagcagcgagcagcagcagcgag 867
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 640 CCGTGTGTGAGAGCTCTGAGGCGCGCGCGGATGCGCAAGTGGCCCTCTACTCTCGAAG 699

QY 868 ggcctcgagcagcgagctgctcagctcgagcagcgagcagcagcagcagcagcagc 927
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 700 GCGCTTGGCGCGCGAGGTTTCTGTTTCG -ACGACGCGGAGACAGNNTCTTCTGACGCC 758

QY 928 gcttcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 944
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 759 GGCCTTGTGTCGACCTTCT 775

```

```

RESULT 2
BF620181
LOCUS
DEFINITION
    BF620181 597 bp mRNA EST 22-FEB-2001
    HVSMEC0018M1f Hordeum vulgare seedling shoot EST library
    HVCNMA0003 (Etiolated and unstressed) Hordeum vulgare cDNA clone
    HVSMEC0018M1f, mRNA sequence.
ACCESSION
    BF620181
VERSION
    BF620181.2 GI:13109232
KEYWORDS
    EST.
SOURCE
    barley.
    Hordeum vulgare
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
    1 (bases 1 to 597)
REFERENCE
    Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Fritsch,D., Yu
    ,Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo
    ,T., Saski,C., Schwartzbeck,J., Simmons,J., Chol,D.W., Main,D. and
    Wood,T.
    Development of a genetically and physically anchored EST resource
    for barley genomics
    Unpublished (2000)
    On Dec 18, 2000 this sequence version replaced gi:11883915.
COMMENT
    Contact: Wing RA
    Clemson University Genomics Institute
    Clemson University
    100 Jordan Hall, Clemson, SC 29634, USA
    Tel: 864 656 7288
    Fax: 864 656 4293
    Email: rwing@clemson.edu
    Seq primer: AATTAACCTTCATCTAAAGG
    High quality sequence stop: 564.
FEATURES
    source
        1..597
        /organism="Hordeum vulgare"
        /cultivar="Morex"
        /db_xref="taxon:4513"
        /clone="HVSMEC0018M1f"
        /clone_lib="Hordeum vulgare seedling shoot EST library
        HVCNMA0003 (Etiolated and unstressed)"
        /tissue_type="Seedling shoot"
        /lab_host="TUC121"
        /note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI. For
        more details on library preparation and sequence analysis
        see http://www.genome.clemson.edu/projects/barley/ To
        order a clone see http://www.genome.clemson.edu/orders"
BASE COUNT      99 a      202 c      204 g      91 t      1 others
ORIGIN

```

```

Query Match      24.3%: Score 516.4; DB 11; Length 597;
Best Local Similarity 92.5%; Pred. No. 7.3e-62;
Matches 554; Conservative 0; Mismatches 42; Indels 3; Gaps 1;

QY 1298 gggaggaagacccgaagagagccgaggaatgagcagctgacagctcttcgagatgc 1357
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 GCGAGGAGGACCTTAACAGAGGCCGAGGATATGCGGTGATCACTTCTCGAGATGC 60

QY 1358 accgagctgctcgagcagcgccctgaggaagctcctgagcagctgagcagctgagc 1417
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 ACCGCTCTCTCGCGAGCGCGCGCCCTGTGAAGTCTCTGGCGACGGTGGCGCGCGTGC 120

QY 1418 ggcacagatcgtaacgctggtgagcagagcgagcagcagcagcagcagcagcagc 1477
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 GCGCATGATGCTACCGCTGTGTGAGAGGAGGCGAGGAGGAGGAGGAGGAGGAGGAG 180

QY 1478 accgcttacagagctctgagcagcagcagcagcagcagcagcagcagcagcagcag 1537
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 ACCGCTTACCGAGTCCCTGTGCACTACTCTCCACCATCTTGCAATCTCTCGAGGGGCA 240

QY 1538 gtcgagcgagcgccacatcgagctcatcgagggtgctgctctcgtcgagcgagcg 1597

```

[illegible]

LOCUS	DEFINITION	FEATURES	SEQUENCE
BE254247	851 bp HVSME0003G24f Hordeum vulgare seedling root EST library HVCDDN0007 (etiolated and unstressed) Hordeum vulgare cDNA clone HVSME0003G24f, mRNA sequence.	EST 23-FEB-2001	

ACCESSION	BF254247
VERSION	BF254247.2
KEYWORDS	GI:13117026 EST.

ORGANISM  
Hordium vulgare  
Eukaryote: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:  
Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae: Pooidae  
1. Triticeae: Hordeum.  
1 (bases 1 to 851)  
REFERENCE  
AUTHORS  
Wing, R., Close, T. J., Kleinhofs, A., Wise, R., Begum, D., Fritsch, D., Yu,  
X., Anderson, H., Dale, J., Henry, D., Kennel, S., Palmer, M., Rambo,  
T., Saski, C., Schwartzbeck, J., Simmons, J., Choi, D. W., Main, D. and  
Wood, T.

TITLE	Development of a genetically and physically anchored EST resource for barley genomics
JOURNAL	Unpublished (2000)
COMMENT	On Nov 16, 2000 this sequence version replaced gi:111833352.

JOURNAL  
COMMENT  
Unpublished (2000)  
On Nov 16, 2000 this sequence version replaced g1:11183352.  
Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: [twing@clemson.edu](mailto:twing@clemson.edu)  
Seq primer: AATTAACCTCCTACTTAAGG  
High quality sequence start: 27  
High quality sequence stop: 611.  
Location/Qualifiers

FEATURES	Location/Qualifiers
source	1. .851

```

/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEF0003G24f"
/clone.lib="Hordeum vulgare seedling root EST library
HVCdNA0007 (etiolated and unstressed)"

```

```

/note=Vector: lambdaZAP, Site_1: EcoRI, Site_2: XhoI. For
more details on library preparation and sequence analysis
see http://www.genome.clemson.edu/projects/parley/ To
order a clone see http://www.genome.clemson.edu/orders"

```

Query Match	23.7%	Score 503;	DB 11;	Length 851;
Best Local Similarity	82.4%;	Pred. No. 5.1e-60;		
Matches 642;	Conservative 0;	Mismatches 120;	Indels 17;	Gaps 5;

QY	1303	gaacacccgaagacaggaagagcccgagatcaatccgcgtcaactaaagctcttgcgaagtgaacagc	1365
Db	2	GGAGAGATCCCCCGAGGAGGCCGAGGTAATCGCCGTGAATCTCAAGTCTTGAGATGACAGCG	61
QY	1363	ctgctcgcgcaagcccgcgccctcgagaaagctctcggcaccgttcgcgcgctgcgcgc	1422
Db	62	CTCTCGCGGAGCGCGCGCGCCCTCGAGAAAGGTCTGTGGAGGTCGCGCGCGTGGCGGCG	121
QY	1423	aggatcgtaacgcttgctggaagcagagcgcaatcacaaatccgcgcacattccctggaacgc	1482
Db	122	AGGATCGTAATCCCCGCTCGAGCAAGAGGCGGAACCAACTCCGGGTCATCTCTGAGACGCG	181
QY	1483	ttacccgaagctctgactactactccacaatggttcgatctccctctgagagcgagaccc	1544
Db	182	TTTCAACCGAGTCCCTGGACATCACTCACTCAACATGTCGATTCCTCTGAGGCGGCAAGCTC	240
QY	1543	ggcgcgcccgccatccgaagctcatcccgggggtctgctgctgcctcgcgcgcgcgcgcgc	1602
Db	241	--GGAGAGCCCGCTCGAGGTCTATGTGGGGGGTGGCGGTCCTGCGCGACCGCGGACG	298
QY	1603	gaaccagtgcatctcgagagtgttaacctcgccgcgcgaatctgcaacgttgctgcgtgcag	1662
Db	299	GACCAGGTATGTCGAGAGGTACTCTCGGCGGCAAGATTCGCAAGTGATGGATGGCTCGAG	358
QY	1663	ggggcgagagcgacacaaagcgccacgcgaagcgtcggcgacagtgcgaagacccgctctggcaac	1722
Db	359	GGCACGTGAGGGCAACAAAGCGGCCCGAGAACCTGGGGCAATGGCGGAACCGCGTGGGCAC	418
QY	1723	ggcggagtgcgaagcgttccacctggggtccaaatgccctacaagcaagcggaacgctgtctg	1782
Db	419	GCGGCGTTCAGAACGCTGCACCTGGGGCTCCATAGCTTAC-AACAGGCGAGCACGCTGCTG	477
QY	1783	gcgcctctcgccgc	1842
Db	478	GCCCTCTTCCGCGGCGGCGACTGCTGCTACGTGAGAGTGAAGAGGAGGCGGCGCTGACTCTC	537
QY	1843	gggtggaacacgc	1902
Db	538	GGGTGGACACGGGCGCCCGCGATCGGCACTTCCGATGGCGGCTTGTGCGGGCGGGA--T	595
QY	1903	cgcgagctttgaacgcgtgttaagtaacatcgtgtag-----catggaagacaacac	1952
Db	596	CGGGAGGTATGAACGCTGTGGGAAGCTTGTGTGAAGGCTGGAGGCGTACACTCAACCAACC	655
QY	1953	agccccgc	2012
Db	656	CGGGCGGCTCCCGTCCCGGCTTGTGGGGCCCGCCACCCCTTGTGTGAGACAAGAAAT	715
QY	2013	gaagctc-aatgtcattgtcagtgtagcgctggaattggaagcagccgctacatcgtctgg	2070
Db	716	GGATTAAGAAGCTAATGTGGGGAAGTCTCTGATGTGGTAACCCGGGTAGACTGATGGG	774

RESULT	4
LOCUS	BF588097
DEFINITION	BF588097 563 bp mRNA EST 12-DEC-2000
ACCESSION	F01_37_F08_b1_A003
VERSION	BF588097
KEYWORDS	propinquinum cDNA, mRNA sequence.
SOURCE	BF588097.1 GI:11680421
ORGANISM	EST. Sorghum propinquinum. Sorghum propinquinum. Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae: PACC Clade: Panicoideae: Andropogoneae: Sorghum. 1 (bases 1 to 563)
REFERENCE	

AUTHORS	Cordonnier-Pratt, M.-M., Gingle, A., Sudman, M., Marsala, C. and Pratt, L.H.
TITLE	An EST database from Sorghum: floral-induced meristems
JOURNAL	Unpublished (2000)
COMMENT	Contact: Cordonnier-Pratt MM

Cordonnier-Pratt, M.-M., Gingle, A., Sudman, M., Marsala, C. and Pratt, L.H.  
An EST database from *Sorghum*: floral-induced meristems  
Unpublished (2000)  
Contact: Cordonnier-Pratt MM  
Department of Botany  
The University of Georgia  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 542 1805  
Email: mmp1ratt@uga.edu  
Sequences have been trimmed to exclude PolyA, vector and regions  
below Phred quality 16. The threshold for highest quality sequences  
is 20.  
Seq primer: JEN REV  
High quality sequence stop: 477  
PolyA-No.

FEATURES	Location/Qualifiers
source	1. .563

```

/organism="Sorghum propinquum"
/db xref="taxon:132711"
/clone_1fb="Floral-Induced Meristem 1 (FM1)"
/note="Organ: Floral-Induced Meristem; Vector:
pLuscript II from Lambda Zap II; Site_1: XhoI; Site_2:
EcoRI; mature plants were placed in a growth chamber for
15 days with 16 hr darkness and 8 hr light (flowering is
induced by short-day conditions); 16 days after being
returned to the greenhouse under natural long days during
late April/early May, meristems were harvested. The
library was made from poly-A RNA in the cloning vector
lambda Zap II. Clones to be sequenced were prepared by
mass excision."

```

BASE COUNT	76 a	217 c	175 g	95 t
ORIGIN				

Query Match	21.88;	Score 464.2;	DB 11;	Length 563;
Best Local Similarity	90.3%;	Pred. No. 1e-54;		
Matches 496;	Conservative	0;	Mismatches 53;	Indels 0;
			Gaps	0

DY 678 gccgcgagcttgcggtcgtcgttgctcgacacgacgaggagccggatctgcgttgttcacgc 737  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 15 GCCCGGCGTCCGGGTGTGTATGACACGCAGAGCCTGGGATCCGCCCTGTTCACGC 74

[illegible]

0y 798 gaagcagatacccttgctgtagcggtcccaaggcggcgcatgcgaagtgcgccta 857  
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||  
Db 135 CAAGCAGATCCCAATGCTGGCTCGTGCACAGGCGGTGCCATTGCCAAGTCGCCGCTTA 194

**Oy**   858 cctcgcgagcgccctgcgccgcggcttcctcgtccgcgcgaacagtccct 917  
||| ||| ||| ||| ||| | ||| ||| |||  
**Dd** 195 ctTTCGGCAGCGCTTGCCCCGCGGTATCCCTTCCC GCCGCACGCCA CACTTT CCT 254

918 cctcgaagccgccttcgcgcgaactccacacgcgcactctagagtcctgcctactact 977  
 |||||  
 255 ccttcgacgcccgccttcgcgcgaactccacacgcgcactctacgagtcctgccttaccct 314  
 |||||

Oy 978 caagttcgcgcacttcacccgccaccaggcatctggagcggttcgcggtccgcg 103  
||||||| ||||||| ||||||| |||| |  
Db 315 CAAGTTCCGCCACTTTCACCGGGAACCAGGCATCTCTGAGGCTTTTCGGC GGTGCGGCG 374

DY 1038 cggcagcgtctcgacttcgcatcaagcagggaatgcagltgccgcacattccagc 109  
||| |||||||||||||||||||||||||||||||||||||||  
Db 375 CGTCCAGCGTCGTGACTTGCGCATCAAGCAGGGGCTGCAGTGGCGGGCTTTCTCCAGGC 434

Qy 1098 cctcgcctctcgtcccgcgagcctccctcgtctcgcctcacgagcgtcgcccccgca 115  
 |||||  
 Db 435 cctcgcctctcgtcccgcgagcctccctcgtctcgcctcacgagcgtcgcccccgca 494  
 |||||

QY 1158 gccgagcagagaccgaacgcctcgcagcaggtggtgtgaaagctcgccagttcgcgcacac 121

Db	QY	Db
495	1218	555
GCAGACGAGACCCACGCGCTTGCACAAGTGGCTGGAAACTGCTCAGTTCCGGGCACAC	catcgcgct 1226	CATCGGCGT 563

RESULT	5
LOCUS	AI881894
DEFINITION	AI881894 593 bp mRNA EST 02-FEB-2000 606074G08.y1 606 - Ear tissue cDNA library from Schmidt lab Ze mays cDNA, mRNA sequence. 606074G08.y1 606

ACCESSION	A1881894	
VERSION	A1881894.1	GI:5566983
KEYWORDS	EST.	
SOURCE	Zea mays.	

ORGANISM  
Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Zea.  
1 (bases 1 to 393)

AUTHORS	Walbot,V.
TITLE	Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL	Unpublished (1999)
COMMENT	Contact: Walbot V

Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2237  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 606074 row: 5 column: 08.

FEATURES	Location/qualifiers
source	1. .593

```

/organism="tea_mays"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/clone_lib="606 - Ear tissue cDNA library from Schmidt

```

```
lab "  
  /tissue_type="mixed"  
  /dev_stage="ear length from 0.5 cm - 2.0 cm"  
  /lab_host="XLOLR (StrataGene)"
```

```

/note="Organ: immature ear; Vector: PBK-CMV; Site_1: EcoRI;
; Site_2: XhoI; Mixed ear tissue cDNA library from Schmidt
lab"

```

BASE COUNT	93 a	210 c	208 g	82 t
ORIGIN				

Query Match	20.9%	Score 445;	DB 10;	Length 593;
Best Local Similarity	85.2%	Pred. No. 4	3e-52;	
Matches 512;	Conservative 0;	Mismatches 80;	Indels 9;	Gaps 1;

Qy 1226 tcgacttcacgtacgcgcgcctcgtcgccgcacgcctcgcgcactcgtgagccgttcatgc 128

Db 1 TCGACTTCACGTACGCTGCGCTCGCGCCACGCTCGCTGACCTGGAGCCGTTTCATGC 60

[illegible]

Oy	1346	tcttcgagatgcaccggcgtgctcgccgaaggccgcgccctygagaagtctcttggcaccg	1400
Dd	121	TGTCGGAGCTGCACCGGTGCTCCTGCCGACCCTTACTCTGCACAAGTCTCTGGGCACCG	180

QY	1406	tgcgcgcgcctgtcgcgccagatcgtccacgtgtgtgagcagagagcgatcaccaactccg	1416
Db	181	TGCCCCGGGTGCGGCCGAGGATCTGACGGTGTGTGGAGCAGGAGGCCAACCACTCCG	240

Qy 1466 gcacattctctgacgcgtctcacaccgagtcctctgtaactactactccacacatgttcgattccc 152  
||||||| ||||||| ||||| ||||||| ||||||| ||||||| |||||

[illegible][illegible]

RESULT	7
LOCUS	BF655617
DEFINITION	BF655617 443 bp mRNA EST 20-DEC-2000
ACCESSION	FM1_46.D10.b1.A003 floral-induced Meristem 1 (FM1) Sorghum
VERSION	propinquum cDNA, mRNA sequence.
KEYWORDS	BF655617
SOURCE	BF655617.1 GI:11920749
ORGANISM	EST.
REFERENCE	Sorghum propinquum.
AUTHORS	Sorghum propinquum.
TITLE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
JOURNAL	Plantae; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
COMMENT	clade; Panicoidae; Andropogoneae; Sorghum.
	1 (bases 1 to 443)
	Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
	,L.H.
	An EST database from Sorghum: floral-induced meristems
	Unpublished (2000)
	Contact: Cordonnier-Pratt MM
	Department of Botany
	The University of Georgia
	Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
	Tel: 706 542 1860
	Fax: 706 542 1805
	Email: mmpratt@uga.edu
	Sequences have been trimmed to exclude PolyA, vector and regions
	below Phred quality 16. The threshold for highest quality sequence
	is 20.
	Seq primer: JEN REV
	High quality sequence stop: 395
	POLYA-No.





[illegible]

QY	1303	gaagaccgcaagagagagccgaggttaatcgcgctcaactaagtccttcagatgacggc	1362
Db	359	GACGACAAAGATGAGAGACCCGAGGTGATCCGCGCACTCAGTGTGAGACTGTGATCGG	418
OY	1363	ctgtcgcgagcccgagcgccc	1384
Db	419	CTGCTGCTCAGCCCGGCGCC	440
RESULT	10		
LOCUS	BE14891	400 bp	mRNA
DEFINITION	MW1001.H03F990430	HTCC MML Wheat Root Library	Triticum aestivum
ACCESSION	BE14891		
VERSION	BE14891.1	GI:9412639	
KEYWORDS	EST.		
SOURCE	bread wheat.		
ORGANISM	Triticum aestivum		
REFERENCE	1 (bases 1 to 400)		
AUTHORS	Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T., Cloutier, S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A., Gustafson, P., Herrmann, R.G., Holton, T., Jacquemlin, J.M., Jia, J., Joudrier, P., Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P., Ogihara, Y., Pecchioli, N., Quiset, C., Schuch, W., Selvaraj, G., Shariflou, M., Sorrells, M., Warburton, M. and Wenzel, G.		
TITLE	International Triticaceae EST Cooperative (ITREC): Production of Expressed Sequence Tags for Species of the Triticaceae		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Warburton M Applied Biotechnology Center, CIMMYT Apdo. Postal 6-641, 06600 Mexico DF MEXICO Tel: 52-5-7269091 ext 1381 Fax: 52-5-7267558/59 Email: mwarburton@cgmnet.com International Triticaceae EST Cooperative (ITREC) http://wheat.pw.usda.gov/genome. Location/Qualifiers 1. 400 /organism="Triticum aestivum" /cultivar="Atlas" /db_xref="taxon:4565" /clone="MW1001.H03" /clone_1lb="ITREC MML Wheat Root Library" /tissue_type="root" /dev_stage="8 day old" /note="Vector: pYES2 (Invitrogen); 0.5-1.5 kbp average insert size."		
FEATURES	source		
BASE COUNT	52 a	165 c	109 g
ORIGIN		67 t	7 others
Query Match	15.8%	Score 335.2;	DB 10;
Best Local Similarity	95.4%	Pred. No. 4.5e-37;	Length 400;
Matches 374;	Conservative 0;	Mismatches 14;	Indels 4;
		Gaps 3;	
OY	889	cgcttcgcgcccgcagcgcgac-agctccctcctctgacgcgcgcttcgcgcgaactcctcca	947
Db	1	CGCTTCGCGNCGGACCGAGCAAGATCCCTCTCTGACGCGCGCTTCGCGGACGCTCTCTCA	60
OY	948	cgcgcaacttcaagaagtcctgcacctcaactcaagtctgcgaacttcaacgcgcaacaaaggc	1007
Db	61	CGCGCACTTCAACGAGTCTCTGCGCCCTACTCCTCAAGTTCCGCCACCTTCAACGCAACGAGC	120
OY	1008	catcttgagcgcttcgcgcgcgctgcgcgcgcgctgcaagctgctgacatcgcgcataaagca	1067
Db	121	CATCTTGAGGCGCTTCGCCGCGCTGCCGCCGCGGTGACGACTGTCTGCACTTCGGCATCAAGCA	180

OY		1068	ggagatgcaagtgcacccgaactcttcagaaccttgcctcgcgtlcccgagggccctccc	1127
Db		181	GGGATGCAGTGTGCCCGCCCTTCACAGACCCTTGCGCTCCTCCCGGCGCCCTCC	240
OY		1128	gtccgcctcacaccgcgttcgcgccccgcgcagccggagacgagaccgccttcagcaggt	1187
Db		241	GTTCCG-CTCACCGCGCTCGGCCCCCAGCCGAGCGAGACGACGCTTGACAGAGGT	299
OY		1188	ggcctcggaagctcgcgccagttcgcgcacaacalcgcgcgttgacttcagtagcgccct	1247
Db		300	GGGCTGGAAAGCTCGCCCACTTGGCCACACAATCGCGCGTCACTTCACTTAGCGMGCCT	359
OY	1248	cgtccgcacccacgcgcgcgagaccttgagccgt	1279	
Db	360	CCTTGC--GCACGTTNGMGACTCTGAGCCCT	389	
RESULT	11			
LOCUS	B1417150	481 bp	mRNA	EST 15-AUG-2001
DEFINITION	B1417150	949053G08.y1.949 - Juvenile leaf and shoot cDNA from Steve Moose		
ACCESSION	B1417150	Zea mays cDNA, mRNA sequence.		
VERSION	B1417150.1	GI:15188173		
KEYWORDS	EST.			
SOURCE	Zea mays.			
ORGANISM	Zea mays.			
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
AUTHORS	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC			
TITLE	clade; Panicoideae; Andropogoneae; Zea.			
JOURNAL	Maize ESTs from various cDNA libraries sequenced at Stanford			
COMMENT	University Unpublished (1999) Contact: Walbot V Department of Biological Sciences Stanford University 855 California Ave, Palo Alto, CA 94304, USA Tel: 650 723 2227 Fax: 650 725 8221 Email: walbot@stanford.edu Plate: 949053 row: G column: 08. Location/Qualifiers 1..481 /organism="Zea mays" /cultivar="W64A" /db_xref="taxon:4577" /clone_id="949 - juvenile leaf and shoot cDNA from Steve Moose" /tissue_type="Immature leaf primordium and vegetative meristem" /dev_stage="4 stages from 3-13 days after imbibing" /lab_host="E. coli XL0LR" /note="Organ: juvenile vegetative shoots; Vector: PAD-GAL4-2.1; Site_1: EcoRI, Site_2: XhoI; Equal amounts of total RNA by weight from 4 tissue sources (see below) were pooled, polyA+ RNA isolated, and cDNA synthesized for EcoRI (5') and XhoI (3') directional cloning into lambda Hybridzap vector from Stratagene. Tissue Sources: 1. Whole shoots 3 days after sowing/imbibing in wet soil. 2. Basal 1.5 cm shoots 6 days after sowing - includes yellow portions of developing leaves 1-5, primordia from 6-8, and the vegetative apex. 3. Non-green portions of developin leaves 4-5 and the vegetative apex, including adult leaf primordia, 9 days after sowing. 4. Partially expanded and greening leaves 4-5 at 13 days after sowing."			
FEATURES	source			
BASE COUNT	62 a	185 c	150 g	75 t 9 others
ORIGIN				
Query Match	15.5%	Score:329,	DB 11:	Length 481;

[illegible]

```

/clone="HV_CEA019J17f"
/clone.lib="Hordeum vulgare seedling green leaf EST
library HVCDNA0004 (Tryptiphe infected & control)"
/tissue_type="seedling green leaf"
/lab_host="TJc121"
/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI; For
more details on library preparation and sequence analysis
see http://www.genome.clemson.edu/projects/barley/To
order a clone see http://www.genome.clemson.edu/orders"

```

## BASE COUNT

```

150 a 290 c 276 g 128 t

```

## ORIGIN

```

Query Match      14.6% Score 309.2; DB 11; Length 844;
Best Local Similarity 79.5%; Pred. No. 1.7e-33;
Matches 410; Conservative 0; Mismatches 88; Indels 18; Gaps 3;

QY 20 cgcgcatactgaagcggagatcaccagacgcggagagcgcgctgagcgca 79
   |||||||
DB 138 cGAGATCTGTAAGCCGCAAGTACCAAGAGCGGCGGAGCGGCGGTGATGAGA 197
   |||||||
QY 80 tgggctctccaggagacaagatgatgt---gtcgcgcgcgcgcgagggagag 136
   |||||||
DB 198 TGGGGTCGTGAGGAGCAAGATGATGTCGTCCGCCGAGCGGAGGAGGAGAG 257
   |||||||
QY 137 tggagagctctgctgcgcgcgtcggtacaaagtgcgcgctccagatgagcgatg 196
   |||||||
DB 258 TGGAGAGCTGCTGCGCGCGGCTGCTACAAAGTGCAGGCTCCGACATGCGGAGCTGG 317
   |||||||
QY 197 cgcagaagctgagagagctcgagatgagcatalggagatggcgcgctgagcgcgcg 256
   |||||||
DB 318 CCGCAGAGCTGAGAGAGCTCGAGATGCGCATGAGCATGGG-----CGGGCCTT 365
   |||||||
QY 257 ccccccagacagacgtctgcacacacacacacacacacacacacacacacac 316
   |||||||
DB 366 CCCCCGAGACGCGCTGTGACCCACCTCGACAGGAGACACCGGAGCATCAACCCACCG 425
   |||||||
QY 317 acctgctcttgggtcgagagatgctgctcgagagctcaagcgcgcgccctcc 376
   |||||||
DB 426 ACCTCTCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 485
   |||||||
QY 377 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 433
   |||||||
DB 486 CCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 545
   |||||||
QY 434 acttcgctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 493
   |||||||
DB 546 ACTTGATGATCCCGGCGCTCTGTCGACTACTCCAGAGAGAACTAGCGCCCTGGGCGCCGAT 605
   |||||||
QY 494 cctcccgcgcgcgcgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 529
   |||||||
DB 606 CATTCCTCCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 641
   |||||||

```

## RESULT 13

```

BE595338      527 bp      mRNA      EST      18-AUG-2000
LOCUS        BE595338
DEFINITION   P11_48-G12.b1-A002 Pathogen induced 1 (P11) Sorghum bicolor cDNA,
ACCESSION    BE595338
VERSION      BE595338.1 GI:9850411
KEYWORDS     EST.
SOURCE       sorghum,
ORGANISM     Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoidae; Andropogoneae; Sorghum.
REFERENCE    1 (bases 1 to 527)
AUTHORS      Cordonnier-Pratt, M.-M., Gingle, A., Dean, R., Sudman, M. and Pratt,
              L.H.
TITLE         An EST database from Sorghum: pathogen-induced plants
JOURNAL      Unpublished (2000)
COMMENT      Contact: Cordonnier-Pratt MM

```

```

Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmprratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 524
POLYA-No.

```

## FEATURES

## source

## Location/Qualifiers

```

1..527
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone.lib="Pathogen induced 1 (P11)"
/note="Organ: Anthracnose-infected leaves from
two-week-old sorghum plants 48 hr after inoculation;
vector: pBluescript II from Lambda Zap II; Site_1: XhoI;
Site_2: EcoRI; Two-week-old sorghum plants (BTX 623
cultivar) were infected with pathogen (isolate PRM421 of
Colletotrichum graminicola, which is a sorghum isolate).
RNA was prepared from infected leaves harvested from 45
seedlings 48 hours after inoculation. Note: young
seedlings (2 weeks old) exhibit juvenile resistant
reaction, which is an incompatible interaction. As they
grow older (4 weeks or older), plants resume susceptibility
to anthracnose disease. The library was made from poly-A
RNA in the cloning vector lambda Zap II. Clones to be
sequenced were prepared by mass excision. WARNING: While
most or all ESTs are expected to derive from the host
plant, no effort was made to eliminate ESTs deriving from
the pathogen."

```

## BASE COUNT

```

65 a 197 c 183 g 82 t

```

## ORIGIN

```

Query Match      14.5% Score 307.4; DB 10; Length 527;
Best Local Similarity 78.3%; Pred. No. 2.9e-33;
Matches 397; Conservative 0; Mismatches 101; Indels 9; Gaps 2;

QY 398 cctccacctctccacacgctcagcgagcgcgctactgactccgcgcctcgtcg 457
   |||||||
DB 21 CCACCTTCACAAAGTGGCGCGCGCGCGCGGTGCTGCTGCTGCTGCTGCTGCTG 80
   |||||||
QY 458 actcctcagacatcagcgtcgcgcgagctccctcccgcgcgcgcgcgcgcgcg 517
   |||||||
DB 81 ACTCGTCAGACAGCACTTACGCTTAAAGCCATCCCTTCCGCGGTGCGGCTCGG 140
   |||||||
QY 518 cgcgcgacgtctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 577
   |||||||
DB 141 CCGACCGCTCTCCACGAGACTCCAGCGGAGAGCGGAGCGGAGCGGAGCGGAG 200
   |||||||
QY 578 gcaactcgtctgctatctctctctctctctctctctctctctctctctctct 634
   |||||||
DB 201 GCACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 260
   |||||||
QY 635 tggtagagctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 688
   |||||||
DB 261 TGGTCAAGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 320
   |||||||
QY 689 cggctcgtctgctcagacagcgagagcgcgagctcgtctgctcagcgctgctg 748
   |||||||
DB 321 CGGTGCTGTGATGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 380
   |||||||
QY 749 ggcgagagcgctgagcgagagagagagagagagagagagagagagagagagag 808
   |||||||
DB 381 GCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 440
   |||||||
QY 809 ccttgtagcgcgctcagcgagcgcgctgagcgagagcgcgagagcgcgagag 868
   |||||||
DB 441 CCATGCTGGCTCTGTCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGAG 500
   |||||||

```

[illegible][illegible]

germinated for three days of the cultivar Williams 82. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dI) sequence with a NotI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by digestions with EcoRI and NotI. The cDNA fragments were directionally cloned into the EcoRI-NotI restriction site of the pT73-Pac vector. The ligated cDNA fragments were transformed into DH10B host cells (Gibco BRL). This library was constructed by Dr. Randy Shoemaker."

BASE COUNT 105 a 182 c 137 g 100 t  
ORIGIN

Query Match 12.9%; Score 273.4; DB 10; Length 524;  
Best Local Similarity 72.1%; Pred. No. 1.3e-28;  
Matches 374; Conservative 0; Mismatches 136; Indels 9; Gaps 1;

```
OY 717 cggagctcgagctgctgacagcgctgctgagcgagcgctgacagagagaacct 776
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 14 cggcgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 73
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 777 ctccgacgcgagagcgagcgctgctgctgctgctgctgctgctgctgctgct 836
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 74 GAAGCTGGCCGACGCGCTGCTGCAAGACAGTGGGCAATCTCGCGCTTCAAGCGCGCGC 133
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 837 gatgacgaagtcgcgcgctgctgctgctgctgctgctgctgctgctgctgct 896
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 134 CATGAGGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 193
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 897 cccgacgcgagcagctgctgctgctgctgctgctgctgctgctgctgctgctgct 956
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 194 CCCGGA-----AGAAACCTTCGACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 244
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 957 ctacgagctcctgacccctgctgctgctgctgctgctgctgctgctgctgctgct 1016
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 245 CTACGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 304
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1017 ggcgttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1076
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 305 AGCATTTGGCACTGCTGAGAGAGTCCAGCTCATGATTTGCGCTCAGACAAAGGATGCA 364
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1077 gtgacgcgactctccagagcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1136
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 365 GTGGCTTGCACTCATGCAAGCCCTTGCAATGCGGCCCGCGCGCGCGCTTTCCGCTT 424
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1137 caccgagctcgagcccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1196
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 425 AACCGAATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 484
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1197 gctgcgccagtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1235
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 485 ATTAGCCCAATTGGCCAGACATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 523
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: January 23, 2002, 01:40:26  
Job time: 9510 sec





[illegible]

RESULT 2  
US-09-117-853-3  
Sequence 3, Application US/091178533  
Patent NO. 6307126  
GENERAL INFORMATION:  
APPLICANT: Harberd, Nicholas P  
APPLICANT: Peng, Jiarong  
APPLICANT: Carol, Jintong

```

APPLICANT: Richards, Donald E
TITLE OF INVENTION: Nucleic acid encoding GAI gene of Arabidopsis thaliana
FILE REFERENCE: 620-45
CURRENT APPLICATION NUMBER: US/09/117,853
CURRENT FILING DATE: 1998-08-12
EARLIER APPLICATION NUMBER: PCT/GB97/00390
EARLIER FILING DATE: 1997-02-12
EARLIER APPLICATION NUMBER: GB 9602796.6
EARLIER FILING DATE: 1996-02-12
NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 1643
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-117-853-3

```

Query Match	16.6%	Score 351.8;	DB 4;	Length 1643;
Best Local Similarity	59.3%;	Pred. No. 3.8e-50;		
Matches 701; Conservative	0;	Mismatches 407;	Indels 75;	Gaps 3

[illegible]



[illegible]

```

US-09-117-853-9
: Sequence 9, Application US/09117853
: Patent No. 6307126
: GENERAL INFORMATION:
: APPLICANT: Harberd, Nicholas P
: APPLICANT: Peng, Jintong
: APPLICANT: Carol, Pierre
: APPLICANT: Richards, Donald E
: TITLE OF INVENTION: Nucleic acid encoding GAI gene of Arabidopsis thaliana
: FILE REFERENCE: 620-45
: CURRENT APPLICATION NUMBER: US/09/117, 853
: CURRENT FILING DATE: 1998-08-12
: EARLIER APPLICATION NUMBER: PCT/GB97/00390
: EARLIER FILING DATE: 1997-02-12
: EARLIER APPLICATION NUMBER: GB 9602796.6
: EARLIER FILING DATE: 1996-02-12
: NUMBER OF SEQ ID NOS: 12
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 9
: LENGTH: 1642
: TYPE: DNA
: ORGANISM: Arabidopsis thaliana
: US-09-117-853-9

```

Query Match	16.3%;	Score 347;	DB 4;	Length 1642;
Best Local Similarity	59.6%;	Pred. No. 2.4e-49;		
Matches 686;	Conservative	0;	Mismatches 390;	Indels 75; Gaps 3

[illegible][illegible]

```

RESULT      4
US-09-117-853-5
; Sequence 5, Application US/09117853
; Patent No. 6307126
; GENERAL INFORMATION:

```



[illegible]

```

Sequence 7 Application US/08804227C
Patent No. 5876991
GENERAL INFORMATION:
APPLICANT: Dehoff, Bradley S.
APPLICANT: Kuhstoss, Stuart A.
APPLICANT: Rostock, Paul R., Jr.
APPLICANT: Sutton, Kimberly L.
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESSES:
ADDRESS: THOMAS G. PLANT 1501
STREET: Lilly Corporate Center
City: Indianapolis
STATE: IN
Country: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435.
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas, G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 44377 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 350..14002
FEATURE:
NAME/KEY: CDS
LOCATION: 14046..20036
FEATURE:
NAME/KEY: CDS
LOCATION: 20110..31284
FEATURE:
NAME/KEY: CDS
LOCATION: 31329..36071
FEATURE:
NAME/KEY: CDS
LOCATION: 36155..41830
US-08-804-227C-7
Query Match 4.7%; Score 99.2; DB 2; Length 44377;
Best Local Similarity 45.4%; Pred. No. 1.4e-08;
Matches 641; Conservative 0; Mismatches 748; Indels 23; Gaps 7;
Qy 240 cgttgagcgccgagcgcgcccccgaacgacagtctgcaccacaactcgcacgagacacct 299
Db 22959 CGGTGACGAGCGCCCGGAGGCCGGAGACCGGCTCGGTAACCACTGCCTGTGAAGGACT 23018
Qy 300 gacatacaaccacacgagacgttgttcttggttcgagaagcatgtctgtcggagctaagc 359
Db 23019 GCGETCA--CCACC GGCGGTGGCGGCCCGGCTGCGCTGTGATCGTCCACCGG 23075
Qy 360 gccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 419
Db 23076 GGACGAGTACACCGCGCTTCGCCGACACCTTGTAACAGGCGGTGCGCTTCCTTCGGCGAAC 23135
Qy 420 ggacgagcgagctacttcgatctccgcgcgcctcgcgtcgcgactctccaaagatctaacgc 479

```

[illegible]

Db	24213	CCCCAGCAGGAGGGTGC-----GGACCCCTTGCCGCGATGCTGGAGCGGCGACACGCGTGGC	24268
Oy	1547	gagggcccaatcgaaatctcatatgtagggagctcgtcgtcgtctctctgcccggcgagcggaagc	1606
Db	24269	GTCGCGTGGCCGACGCTCGATGTGAAGCCGGTTGGCCGCTCTCAACACTGGTGGCTCCG	24328
Oy	1607	aggtcatgtccgaggtgttaacctcgccggcag	1638
Db	24329	CCGTACTCTTGCACGACATCCCGGACGTGCAG	24360

RESULT 7  
US-08-804-198-1

```

; Sequence 1, Application US/08804198
; Patent No. 5945320

```

```

; GENERAL INFORMATION:
; APPLICANT: Burgett, Stanley G.
;

```

APPLICANT: Kuhstoss, Stuart A.  
APPLICANT: Rao, Nagaraaja R.  
APPLICANT: ;

APPLICANT: Richardson, Mark A.  
APPLICANT: Rosteck, Paul R., Jr.  
APPLICANT: Rosteck, Paul R., Jr.

; TITLE OF INVENTION: PLATENOLIDE  
 ; NUMBER OF SEQUENCES: 6  
 ; CORRESPONDENCE ADDRESS:

;;  
CORRESPONDENCE ADDRESS:  
;;  
ADDRESSEE: PAUL R. CANTRELL  
;;  
STREET. ILLIY CORPORATE CENTER  
;

STATE: IN  
CITY: INDIANAPOLIS  
STREET: EIGHT CONGRESS CENTRAL

COUNTRY: USA  
ZIP: 46285

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
;

```

```

;
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh 7
;

```

```

; SOFTWARE: Microsoft Word 5.1
; CURRENT APPLICATION DATA:

```

; APPLICATION NUMBER: US/08/804  
 ; FILING DATE:

;  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
;

NAME: CANTRELL, PAUL R.  
REGISTRATION NUMBER: 36,470

REFERENCE/DOCKET NUMBER: P911  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317 276 2005

TELEPHONE: 31/-276-3885  
; INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:

```

; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 44377 base pairs
; TYPE: nucleic acid

```

```

; LIFE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

MOLECULE TYPE: DNA (genomic)

NAME/KEY:	CDS
LOCATION:	350..14002

```
FEATURE:
NAME/KEY: CDS
```

```

; LOCATION: 14046..20036
;
; FEATURE:

```

```

; NAME/KEY: CDS
; LOCATION: 20110..31284
;

```

```

;
; FEATURE:
; NAME/KEY: CDS
;

```

```

; LOCATION: 31329..36071
; FEATURE:
; NAME/AREA:

```

```

; NAME/KEY: CDS
; LOCATION: 36155..41830
;
TC-08-804-108-1

```

05-08-804-198-1

Query Match	4.7%	Score
Best Local Similarity	45.4%	Prevalence

Matches 641; Conservative 0;

[illegible]

QY	1311	gaagagagagcccgagagtaatcgcgtcaactaactgattcttgagatagacacggtctgtcgc	1370
Db	24033	ggggggccggcccccaggggggggctaaaggccggccggccagccgacactcagacggcgctggccga	24092
QY	1371	gcagcccgagcgccctcgagaaagatctccgagccgtg----cgcgcgctgagcccgaga	1426
Db	24093	ACGCGCCCGCGCCCGGGGGGGGTGCCGCGTTCTCCCTGGCCTGGAGCCCTTGCGCCGGAGG	24152
QY	1427	tcgtacccgtgtgtgagacagagacgaaatcacaaactccgagacatctctgtgacgcttca	1486
Db	24153	CACGCGCCGCCAGAGGTGCGAGAGCGAGATTCTCTACGCGGGCGGGCGTGGCTCCCTCGA	24212
QY	1487	ccgagctctctgactactactaccacatglttcgattccctcagagcgagctccgagc	1546
Db	24213	CCCCAGCAGGGGGTGC---GGACCTTGCGGGCATGCTGGAGGCGGCAAGCCCTCGC	24268
QY	1547	gcggcccatccgaagtctcatcgcggggtgctgtctgtctctgtccgcgccgagcacggaac	1606
Db	24269	GTGCGGTGCGCGACGTCAAGTAGAGACCGGTTCCGCGCTCTACACCTGGGTGCTTCCG	24328
QY	1607	aggtcatgtccgaggtgtacctcgcgcgag	1638
Db	24329	CCGTACTCTTGACAGACATCCCGGACGTGGAC	24360

```

RESULT      8
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

[illegible]

```

Db 3929228 CC--GCTCCGCGCGCCCGCTTGATGCGCGCGCGCGCTTGCGCGCGCGCTT 3929171
QY 533 ccgactccgltgagagatcccaagagagatgacatgagagagagacactgctgcat 592
Db 3929170 GCGCGCTTTGGCGCGCGCGCGAGAGCGCTTGTGGAGAGAGCGCGCGCTTGGCGG 3929111
QY 593 cctccctccctgctgctcctcgttggagagagagctctgtgtgagagctgccccg 652
Db 3929110 CCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGATACCGCGCTTGGCGCTTCCG 3929051
QY 653 cgtgctgagcgcgcgagagagagagagagagagagagagagagagagagagag 712
Db 3929050 CGGTGTGGCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCTGCGCGCGCGCGCGCGAGC 3928991
QY 713 agagcgagagatcgctgctgctgctgctgctgctgctgctgctgctgctgctg 772
Db 3928990 CGCGCATTAAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTGC 3928931
QY 773 acctctccgcgagagagagagagagagagagagagagagagagagagagagag 832
Db 3928930 CGCGCTACACCGCTGCGCGCGCGCGCTGCGCTGCGCTGCTT--CAGGCGCGCTTGGCGGCT 3928872
QY 833 gcgcgagatgcgcaagctgcgcgctactctgcgcgagagagagagagagagagag 892
Db 3928871 TGGCGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3928812
QY 893 tcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 952
Db 3928811 CCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGAGACA 3928752
QY 953 actctcaagatctcgtccctcaactcaactcaactcaactcaactcaactcaactca 1012
Db 3928751 CCGGCGCGCGAGATCGCGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3928692
QY 1013 tggagagcgltgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1072
Db 3928691 CCGTGGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3928632
QY 1073 tgcagtgagcgcaactctcctcagagagagagagagagagagagagagagagag 1132
Db 3928631 CCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3928572
QY 1133 gctcaacgcgcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1192
Db 3928571 CCGCGCTTGGCGCGCGCGCGAGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCG 3928512
QY 1193 ggaagctgcgcagctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1252
Db 3928511 CCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3928452
QY 1253 ccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1312
Db 3928451 GTGTGGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3928392
QY 1313 acgagagagccgagagatgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1372
Db 3928391 CCGATTAAGCGG-----TGGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3928338
QY 1373 agccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1432
Db 3928337 TTGGCGCGCGTACCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTGC 3928278
QY 1433 ccgtgtgtgagagagagagagagagagagagagagagagagagagagagagagag 1492
Db 3928277 CCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3928218
QY 1493 cctctgactactactcaacatgctcgtcctcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1552
Db 3928217 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3928158
QY 1553 catcgagatcactcagggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1612

```

```

Db 3928157 ACACCGCGCGCGAATCCGGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGAGACA 3928098
QY 1613 tgcgcaggtgtactcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1672
Db 3928097 AATCCGGGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3928038
QY 1673 gacagagagcgcaagagagagagagagagagagagagagagagagagagagagag 1732
Db 3928037 CCGCGGATCAGAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3927978
QY 1733 agacgcgtcactggtgtcctcaatgtcctcaagcagagagagagagagagagagag 1792
Db 3927977 CCGGCGCGCGTGAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3927918
QY 1793 ccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1824
Db 3927917 CCGCGCGTAAAGCGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3927886

```

```

RESULT 9
US-09-130-114-2
; Sequence 2, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: HORLICK, Robert A.
; APPLICANT: DAMEJ, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Strably Expressing Genes
; FILE REFERENCE: 0867/ID903051
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1931
; TYPE: DNA
; ORGANISM: EBNA
US-09-130-114-2

```

```

Query Match 4.3%; Score 91.8; DB 2; Length 1931;
Best Local Similarity 45.9%; Pred. NO. 1.9e-07;
Matches 351; Conservative 0; Mismatches 412; Indels 2; Gaps 1;
QY 806 tacctctgtgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 865
Db 321 tcccgctcctcgtctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 380
QY 866 aggcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 925
Db 381 tctcctcgcgtcgtcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 440
QY 926 ccgcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 985
Db 441 ccgtcctcgtcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 500
QY 986 cgcactcaacgcgaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1045
Db 501 tcccgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 560
QY 1046 tctgtactcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1105
Db 561 tctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 620
QY 1106 tccgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1165
Db 621 tctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 680
QY 1166 agacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1225
Db 681 tctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 740

```

QY	1226	tcgattccagatcacgcggcctcgtctgcgcgcacagctcgcgcgacctgtgagccgttcatatgc	1285
Db	741	tctctcccgctctctcccgctctcgtctctctcccgctctctcccgctctctctcccgctcc	800
QY	1286	tgcagcgcggagggcgagaggaagccgaagagagcgccaggttaatcgcgcgtcaactcag	1345
Db	801	tccccgctctctcgtctctctctcccgctctcccgctctctcgtctctctcccgctctcccg	860
QY	1346	tcttcagatgcagacgcctcgtctgcgcgcgcgcgcgcctgtgagaagctctcttgcgcacgc	1405
Db	861	tctctcgtctctcccgctctctcccgctctctcgtctctctcccgctctcccgctctctctctcc	920
QY	1406	tgcgcgcgcgtgcgcgcgcacagatcgttcaaccgttggctggagcagagagcgatcacactcgc	1465
Db	921	ccgtctctcccgctctcgtctctctctctcccgctctctctgc - tcccgctctcgtctctcca	978
QY	1466	gcacattctctgcagcgccttcaaccgagctctctgactactactcaacacatgttcgattccc	1525
Db	979	ctctcgcgcgcccaatctctctcgttcaactctcgcgcgcacagctctctcactcaactcgcgc	1038
QY	1526	tgcagggcggcagctcgcgcgcgcgcgcgcacatctcgaaagctcacaacgc	1570
Db	1039	gctctctcattcaactctcgcgcgcgcgcacatctctgcgaattctcgc	1083

```

RESULT 10
US-09-105-537-1
Sequence 1, Application US/09105537A
Patent No. 6265202
GENERAL INFORMATION:
APPLICANT: Sheerman, D. H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600.438051
CURRENT FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: fastseq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 15872
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-105-537-1

```

Query Match	4.2%;	Score 89.2;	DB 4;	Length 15872;
Best Local Similarity	43.2%;	Pred. No. 5.7e-07;		
Matches 580;	Conservative 0;	Mismatches 753;	Indels 9;	Gaps 3

[illegible][illegible]

RESULT 11  
US-09-105-537-32













Db 15402 CAGAGGCGCGTGGAGCTCGCCCTCTTTCGAGCTGCTCGCCGAGCGCTCCCGGCGTCCGAC 15461  
QY 268 agcttcgcaaccacactcgccacgagacaccggtgcaactaacaacccagacactgctgctc 327  
Db 15462 CAGATTCGCGCCGGAAGTGGGCGCTCCGGCACCGGCTTTAAAGAGACTCTCTGAACGGCTTG 15521  
QY 328 tgggtcgaagagctgctgctgagagctcaacgagcgccgcccctcccgccgcccgcg 387  
Db 15522 GTCCGGCTCGGCGCTGCTGGACGAGGAGCGGGGGCTTACCGGGGCTCCGAGCTCGGCGCG 15581  
QY 388 cagctcaacgctcccaactcctcccaacgctcaacgagcgagcgagcgagctgactatcccg 447  
Db 15582 GAGTTCCTCTCCCGCGCCCACTGATCTCGGCGGACCGCCCGCAGCACCTCTCGGCTG 15641  
QY 448 ccttcgctgacttcctcccaagacatctcaagcgtgagcgagctcccccctcccgccgccc 507  
Db 15642 CACTTACCACTGCTGGGCGGACGCTGACCCGACGCGCTGCGGACGGCAAGGCTCAAGTGGCC 15701  
QY 508 gcgacgagcgccgagccgacactgctccgacgctccgctgagaga-tcccaagcgatgagcac 566  
Db 15702 GTGGCGCGGCGGACGAGGCGAGCTGGCTTACCCCAAGCACTAGAGAGCTGGACCGCGCGCG 15761  
QY 567 tggcgaggagcagacactgctgctcactcctcctcctcgtcgtcgtggtgggagcgag 626  
Db 15762 CAGATCATGCTGACATGAGCGCCCAAGCGGTTTCACGCGCGACGAGTGGCGCGCGCG 15821  
QY 627 gagctctgctgctgagagcgccgcccggctgagcgagcgagcgaagcgagcgccgctc 686  
Db 15822 ATCGACTGGAGCGCGGACTACCTCTCTGAGAGTGGCGGCGCGCGCGCAACGTCGCC 15881  
QY 687 ggc--ggtcgtcgtggtgcaacgca---ggaagccgagatcgagctgagcgagctg 741  
Db 15882 TCCCGGATGTCACCGGCCACCGGACCTGGCGGGGGGTCTTCGACCTCGCGGGCGCTG 15941  
QY 742 ctgagctgagcgagcgctgcaagagagaacactcctccgagcgagagcgagcgctggtgaag 801  
Db 15942 CCGCGGCTCTTGGAGAGAGCTGGTGGCTCGGCGGAAACCGCGAGCGGCTGAGCTTCAC 16001  
QY 802 cagataacctgctgagcgagctcccaagcgagcgagctgcaaggtcgagcgctcctc 861  
Db 16002 GCGCGTACTTCTTCCGCAACCGACTGCGGAGGCGGACGTGGTATCTTCGTCAGCC 16061  
QY 862 ggcgagggccctcgccgagcgctctcctccgctccgagccgag---acagctccctc 918  
Db 16062 TCCCGGACTGGGCGCGGTCCGGGACCGCAAGAGCTGCTGCGCGCCCAAGGCGGTG 16121  
QY 918 ctgagagcgagctcgagacactcctcccaacgagcgagctcctgagctcctgcccctac 978  
Db 16122 CCGCGGCGGCGCTGGTGGTGTGCTGACGAGCGCATGATGACCGGAGGAGCGGACCGCC 16181  
QY 978 aagttcgagcaactcaacgcaacgaagggatccctggaagcggtcgagcgctgagcgccg 1038  
Db 16182 GAGGTCTCTCTCAGCGGATCAACACCACTGCGGAGCGGAGCGCGGCTTACTCG 16241  
QY 1038 gtcagagctcgagctcgagctcaacgaagggatgagtgagcccgagctcctccagggcc 1098  
Db 16242 CTGCAAGGAGGCGCGGCTTACTGAGGAGGCGCGGCTTCAACGTCGACCGGATCGCGCGC 16301  
QY 1098 ctgagcctcgctccgagcgagcctccctcgttcgagctca--ccgagctcgagcccgccg 1156  
Db 16302 TCCGACACCATCAACCGCGGACCACTTCGCATCGGCTCAAGTCGGTCTGAAGAAAGG 16361  
QY 1157 agcgagagagagagagcgcccgagcagagtgagctgagagctcg---cagttcgagc 1213  
Db 16362 AGTTGACATGACCGGACACACCGCGGATCAACGCGGACCGCGGCTTACGAGC 16421  
QY 1214 aacacacacgctgagctcagctcaacgagcgagctcgagcgagcgagcgagcgagctg 1273  
Db 16422 TCGTATCATTCGGCAAGCGGCTGTGCGGACCATGCTCGGCTGATCTCGCAACGACG 16481  
QY 1274 agcgctcagctgagcgagcgagcgaggaagaccgagagagcccgagagtaatcg 1333

Db 16482 GCTTCCGATCATGCTGCTGAGCGGTGCCACACCCCGCTTCGCGCTGGCGAGTCCA 16541  
QY 1334 ccgtcaactcagctctgagatgcaacggcgctgctcgagcagccgagccctggagaaag 1393  
Db 16542 CCATCGGAGAGAGCGGTGCTGCTGCGGCTGATCTCGGACCGGATGAGGGGTGCGGAGCA 16601  
QY 1394 tccctggacagctgagcgagcgagctgagccagagatcgtaacgctggtgagagagagcga 1453  
Db 16602 TCGCCAACTGGCGAGCTTCCAGGAGCTCTCGCCAACTGACAGATTTCGACGCGGACGA 16661  
QY 1454 atcaaacctccgagacatctcctgagacgcttcacagagctcgtgactactactacaca 1513  
Db 16662 AAGACAACTTGGCTTCAATGTTCCACCGGGAGGAGGAGCGGACCGGAAGAGACCA 16721

Search completed: January 23, 2002, 05:43:16  
Job time: 23626 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2002, 16:25:59 ; Search time 40.57 Seconds

(without alignments)  
23.736 Million cell updates/sec

Title: US-09-485-529-103

Perfect score: 76

Sequence: 1 LNA PPPPLPPAPQ 13

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_1101.\*

1: /SID58/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID58/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID58/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID58/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID58/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID58/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID58/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID58/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID58/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID58/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID58/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID58/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID58/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID58/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID58/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID58/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID58/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID58/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID58/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID58/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID58/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	76	100.0	623	20	AAV02540
2	59	77.6	256	20	AAV02538
3	58	76.3	33	21	AAAG27119
4	57	75.0	182	22	AAAB80893
5	57	75.0	216	22	AAAB80892
6	57	75.0	223	18	AAAW21733
7	57	75.0	282	22	AAAB95839
8	57	75.0	282	22	AAAB80896
9	57	75.0	321	22	AAAB95274
10	56	73.7	123	20	AAV02543
11	56	73.7	630	20	AAV02541

12	55	72.4	293	21	AAAB43674
13	55	72.4	313	22	AAAB79217
14	55	72.4	313	22	AAAB79222
15	55	72.4	347	21	AAAG14025
16	55	72.4	356	21	AAAG14024
17	55	72.4	359	21	AAAG14023
18	55	72.4	438	22	AAAG90541
19	55	72.4	446	12	AAAR1837
20	55	72.4	446	16	AAAR72484
21	55	72.4	448	14	AAAR39467
22	54	71.1	90	21	AAAG19012
23	54	71.1	113	21	AAAG11750
24	54	71.1	138	21	AAAG11749
25	54	71.1	148	21	AAAB42921
26	54	71.1	167	21	AAAG11748
27	54	71.1	148	18	AAAB26545
28	54	71.1	527	20	AAV29515
29	54	71.1	527	21	AAAB44492
30	54	71.1	580	22	AAAB60493
31	54	71.1	2343	21	AAAB12453
32	54	71.1	2453	21	AAAB12454
33	53	69.7	55	20	AAV04937
34	53	69.7	104	22	AAAM15554
35	53	69.7	104	22	AAAM28043
36	53	69.7	104	22	AAAM03303
37	53	69.7	212	21	AAAG36695
38	53	69.7	237	21	AAAG36694
39	53	69.7	257	20	AAV25630
40	53	69.7	270	21	AAAG36693
41	53	69.7	285	20	AAV25633
42	53	69.7	373	20	AAV04940
43	53	69.7	390	21	AAAB58287
44	53	69.7	427	20	AAV04941
45	53	69.7	561	19	AAAB63701

#### ALIGNMENTS

RESULT 1  
ID AAV02540 standard; Protein: 623 AA.  
XX  
AC AAV02540;  
XX  
DT 16-JUL-1999 (first entry)  
XX  
XX Protein encoded by wheat Rht clone 5a1 genomic sequence.  
XX  
KW Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;  
KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;  
KW paclobutrazol.  
XX  
OS Triticum aestivum.  
XX  
PN WO9909174-A1.  
XX  
PD 25-FEB-1999.  
XX  
PF 07-AUG-1998; 98WO-GB02383.  
XX  
PK 13-AUG-1997; 97GB-0017192.  
XX  
PA (PLAN-) PLANT BIOSCIENCE LTD.  
XX  
PI Harberd NP, Peng J, Richards DE;  
XX  
DR WPI: 1999-181040/15.  
XX  
DR N-PSDB; AAX36279.  
XX  
PT New Triticum aestivum polynucleotides - encode a polypeptide which provides inhibition of the growth of plants, which inhibition is antagonised by gibberellin, used to confer a dwarf phenotype  
PT

XX Disclosure; Fig 8b; 88pp; English.  
PS  
XX  
CC The specification describes polypeptides encoded by the Rht gene (and  
CC its homologues) that, when expressed in Triticum Aestivum, inhibit  
CC growth of the plant. This growth inhibition is antagonised by  
CC gibberellin. The products can be used to provide Rht expression in  
CC plants, conferring a dwarf phenotype on a plant which is correctable  
CC by treatment with gibberellin. In addition, the products can be  
CC used to produce Rht mutant plants which are dwarfed compared with  
CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants  
CC may be made by knocking out Rht or the relevant homologous gene in  
CC the plant of interest. Plants may be made which are resistant to  
CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,  
CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds  
CC dwarf but let crop plants grow tall. The present sequence is encoded by  
CC the wheat Rht clone 5a1 genomic sequence.  
CC  
XX  
SQ Sequence 623 AA:  
  
Query Match 100.0%; Score 76; DB 20; Length 623;  
Best Local Similarity 100.0%; Pred. No. 0.27;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 LNAPPPPLPPAPQ 13  
|||  
DB 109 lnappplppapq 121  
  
RESULT 2  
AAV02538  
ID AAV02538 standard; Protein; 256 AA.  
XX  
AC AAV02538;  
XX  
DT 16-JUL-1999 (first entry)  
XX  
DE Protein encoded by rice EST D39460 sequence.  
XX  
KW Rht gene; homologue: Triticum aestivum; wheat; growth inhibition;  
KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;  
KW paclobutrazol; rice; expressed sequence tag; EST.  
XX  
OS Oryza sativa.  
XX  
PM MO9909174-A1.  
XX  
PD 25-FEB-1999.  
XX  
PF 07-AUG-1998; 98MO-GB02383.  
XX  
PR 13-AUG-1997; 97GB-0017192.  
XX  
PA (PLAN-) PLANT BIOSCIENCE LTD.  
XX  
PI Harberd NP, Peng J, Richards DE;  
XX  
DR WPI, 1999-181040/15.  
XX  
DR N-PSDB; AAX36277.  
XX  
PT New Triticum Aestivum polynucleotides - encode a polypeptide which  
PT provides inhibition of the growth of plants, which inhibition is  
PT antagonised by gibberellin, used to confer a dwarf phenotype  
XX  
PS Clatum 12; Fig 6b; 88pp; English.  
XX  
CC The specification describes polypeptides encoded by the Rht gene (and  
CC its homologues) that, when expressed in Triticum Aestivum, inhibit  
CC growth of the plant. This growth inhibition is antagonised by  
CC gibberellin. The products can be used to provide Rht expression in  
CC plants, conferring a dwarf phenotype on a plant which is correctable  
CC by treatment with gibberellin. In addition, the products can be

CC used to produce Rht mutant plants which are dwarfed compared with  
CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants  
CC may be made by knocking out Rht or the relevant homologous gene in  
CC the plant of interest. Plants may be made which are resistant to  
CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,  
CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds  
CC dwarf but let crop plants grow tall. The present sequence is encoded by  
CC wheat Rht gene.  
CC  
XX  
SQ Sequence 256 AA:  
  
Query Match 77.6%; Score 59; DB 20; Length 256;  
Best Local Similarity 83.3%; Pred. No. 9.1;  
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
OY 1 LNAPPPPLPPAP 12  
|||  
DB 109 lnappplppap 120  
  
RESULT 3  
AAG27119  
ID AAG27119 standard; Protein; 33 AA.  
XX  
AC AAG27119;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Zea mays protein fragment SEQ ID NO: 31833.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence; corn.  
XX  
OS Zea mays subsp. mays.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0123788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.



PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135533.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136621.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0160770.  
PR 18-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.

PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 76.3%; Score 58; DB 21; Length 33;

Best Local Similarity 90.0%; Pred. No. 2;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPAPQ 13  
1111111111  
DB 13 PPPPLPPAPR 22

RESULT 4

AAB80893  
ID AAB80893 standard; protein; 182 AA.

AC AAB80893;

DT 30-MAY-2001 (first entry)

DE Human ATFx leucine zipper domain.

KW Human; ATFx; gamma aminobutyric acid B receptor; GABA B receptor;  
transcription factor; leucine zipper.

OS Homo sapiens.

PN WO200116596-A2.

PD 08-MAR-2001.

PF 31-AUG-2000; 2000MO-GB03361.

PR 31-AUG-1999; 99GB-0020569.

PR 12-JAN-2000; 2000GB-0000516.

PA (GLAX ) GLAXO GROUP LTD.

PI White J, Wise A, Marshall F;

DR WPI: 2001-265904/27.

DR N-PSDB; AAF77878.

PT Identifying modulators of gamma aminobutyric acid(GABA)-B receptor  
mediated activity by monitoring the interaction between GABAB receptor  
and the CREB/ATF transcription factors in the presence of a test  
substance

PS Claim 14; Page 64-65; 71pp; English.

CC The present invention relates to a method for identifying a modulator of  
CC gamma aminobutyric acid<sub>B</sub> (GABA<sub>B</sub>) receptor-mediated activity, by  
CC monitoring the interaction between a CREB/ATF transcription factor  
CC capable of binding to GABA<sub>B</sub> receptor. The present sequence is the  
CC leucine zipper domain of human ATFx (see AAB80892), which was used in the  
CC method of the present invention. ATFx is a member of the CREB/ATF family  
CC of transcription factors. ATFx contains a bZIP domain, by which it  
CC interacts with the coiled coil domain of GABA<sub>B</sub> receptor. Modulators of  
CC GABA<sub>B</sub> receptor activity are useful for treating central nervous system  
CC or peripheral nervous system disorders.

XX Sequence 182 AA;

Query Match 75.0%; Score 57; DB 22; Length 182;

Best Local Similarity 100.0%; Pred. No. 11;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPAP 12  
1111111111

DB 31 PPPPLPPAP 39

RESULT 5

AAB80892  
ID AAB80892 standard; protein; 216 AA.

AC AAB80892;

DT 30-MAY-2001 (first entry)

DE Human ATFx.

KW Human; ATFx; gamma aminobutyric acid B receptor; GABA B receptor;  
transcription factor.

OS Homo sapiens.

PN WO200116596-A2.

PD 08-MAR-2001.

PF 31-AUG-2000; 2000MO-GB03361.

PR 31-AUG-1999; 99GB-0020569.

PR 12-JAN-2000; 2000GB-0000516.

PA (GLAX ) GLAXO GROUP LTD.

PI White J, Wise A, Marshall F;

DR WPI: 2001-265904/27.

DR N-PSDB; AAF77877.

PT Identifying modulators of gamma aminobutyric acid(GABA)-B receptor  
mediated activity by monitoring the interaction between GABAB receptor  
and the CREB/ATF transcription factors in the presence of a test  
substance

PS Claim 14; Fig 1; 71pp; English.

CC The present invention relates to a method for identifying a modulator of  
CC gamma aminobutyric acid<sub>B</sub> (GABA<sub>B</sub>) receptor-mediated activity, by  
CC monitoring the interaction between a CREB/ATF transcription factor  
CC capable of binding to GABA<sub>B</sub> receptor. The present sequence is human  
CC ATFx, which was used in the method of the present invention. ATFx is a  
CC member of the CREB/ATF family of transcription factors. ATFx contains a  
CC bZIP domain, by which it interacts with the coiled coil domain of GABA<sub>B</sub>  
CC receptor. Modulators of GABA<sub>B</sub> receptor activity are useful for treating  
CC central nervous system or peripheral nervous system disorders.

XX Sequence 216 AA;

Query Match 75.0%; Score 57; DB 22; Length 216;

Best Local Similarity 100.0%; Pred. No. 13;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPAP 12  
1111111111  
DB 65 PPPPLPPAP 73

RESULT 6

AAW21733  
ID AAW21733 standard; Protein; 223 AA.

AC AAW21733;

DT 01-OCT-1997 (first entry)

DE NIP-1 encoded by clone 59.

KM NIP-1; NIP-2; NuMA; nuclear mitotic apparatus; NuMA interacting protein;  
 KM cell division; proliferation; antibody; Ab; detection;  
 KM malignant cell growth.  
 OS Homo sapiens.  
 XX  
 XX WO9640917-A1.  
 XX  
 PD 19-DEC-1996.  
 XX  
 XX 07-JUN-1996; 96WO-US09504.  
 XX  
 PR 07-JUN-1995; 95US-0478408.  
 XX  
 PA (UYUA ) UNIV YALE.  
 XX  
 PI McPherson SMG, Snyder MP;  
 XX  
 DR MPI; 1997-077270/07.  
 DR N-PSDB; AAT77784.  
 XX  
 XX New nucleic acid encoding nuclear mitotic appts. interacting  
 PT proteins - useful for modulating cell division and proliferation and  
 PR in diagnosis  
 XX  
 PS Claim 28; Page 57-58; 78pp; English.  
 XX  
 CC The sequences given in AAW21733-35 represent NuMA binding proteins (NuMA  
 CC - nuclear mitotic apparatus). These protein sequences were identified  
 CC using the fusion proteins given in AAW21731-32. Compounds which  
 CC interfere with the interaction of NuMA with a known NIP (NuMA interacting  
 CC protein) are used to modulate cell division and/or proliferation. Ab,  
 CC raised conventionally using NIP-1 or -2 as immunogen, are used to detect  
 CC NIP (or their complexes) and to block their activity for diagnostic or  
 CC therapeutic use, e.g. to detect defective NuMA or NIP which may be  
 CC markers for aberrant (including malignant) cell growth (which can also  
 CC be detected by nucleic acid sequencing). Also where malignancy is  
 CC related to defects in NuMA or NIP, it can be treated by administration  
 CC of the appropriate functional protein. This protein is rich in  
 CC proline residues.  
 CC  
 CC Sequence 223 AA:  
 SQ

Query Match 75.0%; Score 57; DB 18; Length 223;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLPPAP 12  
 Db 2 PPPPLPPAP 10  
 DB 2 PPPPLPPAP 10

RESULT 7  
 AAB95859  
 ID AAB95859 standard; Protein; 282 AA.  
 XX  
 AC AAB95859;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human protein sequence SEQ ID NO:18922.  
 XX  
 KM Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-0116126.  
 XX

PR 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR MPI; 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 PS Claim 8; SEQ ID 18922; 2537pp + CD ROM; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides, and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 CC  
 CC Sequence 282 AA:  
 SQ

Query Match 75.0%; Score 57; DB 22; Length 282;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLPPAP 12  
 Db 131 PPPPLPPAP 139  
 DB 131 PPPPLPPAP 139

RESULT 8  
 AAB80896  
 ID AAB80896 standard; Protein; 282 AA.  
 XX  
 AC AAB80896;  
 XX  
 DT 30-MAY-2001 (first entry)  
 XX  
 DE Human ATFX with N-terminal extension.  
 XX  
 KM Human; ATFX; gamma aminobutyric acid B receptor; GABA B receptor;  
 KM transcription factor.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200116596-A2.  
 XX  
 PD 08-MAR-2001.  
 XX

XX 31-AUG-2000; 2000WO-GB03361.  
 PF 31-AUG-1999; 99GB-0020569.  
 PR 12-JAN-2000; 2000GB-0000510.  
 PR (GLAXO ) GLAXO GROUP LTD.  
 PA  
 XX White J, Wise A, Marshall F;  
 PI  
 DR WPI: 2001-265904/27.  
 DR N-PSDB; AAF77880.  
 XX  
 PR Identifying modulators of gamma aminobutyric acid(GABA)-B receptor  
 PT mediated activity by monitoring the interaction between GABAB receptor  
 PR and the CREB/ATF transcription factors in the presence of a test  
 PR substance -  
 PS  
 XX Claim 14; Fig 9; 71pp; English.  
 PS  
 CC The present invention relates to a method for identifying a modulator of  
 CC gamma aminobutyric acid\_B (GABA\_B) receptor-mediated activity, by  
 CC monitoring the interaction between a CREB/ATF transcription factor  
 CC capable of binding to GABA\_B receptor. The present sequence is human  
 CC ATfx, which was used in the method of the present invention. ATfx is a  
 CC member of the CREB/ATF family of transcription factors. ATfx contains a  
 CC bzip domain, by which it interacts with the coiled coil domain of GABA\_B  
 CC receptor. Modulators of GABA\_B receptor activity are useful for treating  
 CC central nervous system or peripheral nervous system disorders. The  
 CC present sequence has an N-terminal extension compared to AAB80892.  
 CC  
 SQ Sequence 282 AA;

Query Match 75.0%; Score 57; DB 22; Length 282;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 4 PPPPLPPAP 12  
 |||||  
 Db 131 PPPPLPPAP 139

RESULT 9  
 AAB95274  
 ID AAB95274 standard; Protein: 321 AA.  
 XX  
 AC AAB95274;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human protein sequence SEQ ID NO:17476.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 OS Homo sapiens.  
 XX  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-0116126.  
 PF  
 PR 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI: 2001-318749/34.  
 DR  
 XX  
 PR Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PR and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 PS Claim 8; SEQ ID 17476; 2537pp + CD ROM; English.  
 PS

CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 CC  
 SQ Sequence 321 AA;

Query Match 75.0%; Score 57; DB 22; Length 321;  
 Best Local Similarity 76.9%; Pred. No. 18;  
 Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Oy 1 LNAPPPPLPPAPQ 13  
 | |||| |  
 Db 121 LPppppppppapq 133

RESULT 10  
 AA02543  
 ID AA02543 standard; Protein: 123 AA.  
 XX  
 AC AA02543;  
 XX  
 DT 16-JUL-1999 (first entry)  
 XX  
 DE Protein encoded by the partial sequence of the maize D8-2023 allele.  
 XX  
 KW Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;  
 KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;  
 KW paclobutrazol; maize.  
 XX  
 OS Zea mays.  
 XX  
 PN WO9909174-A1.  
 PN  
 PD 25-FEB-1999.  
 PD  
 PF 07-AUG-1998; 98WO-GB02383.  
 PF  
 PR 13-AUG-1997; 97GB-0017192.  
 PR  
 XX (PLAN-) PLANT BIOSCIENCE LTD.  
 XX  
 PI Harberd NP, Peng J, Richards DE;  
 XX

DR WPI: 1999-181040/15.  
 DR N-PSDB; AAX36282.  
 XX  
 PT New Triticum aestivum polynucleotides - encode a polypeptide which  
 PT provides inhibition of the growth of plants, which inhibition is  
 PT antagonised by gibberellin, used to confer a dwarf phenotype  
 XX  
 PS Disclosure: Fig 11d; 88pp; English.  
 CC The specification describes polypeptides encoded by the Rht gene (and  
 CC its homologues) that, when expressed in Triticum aestivum, inhibit  
 CC growth of the plant. This growth inhibition is antagonised by  
 CC gibberellin. The products can be used to provide Rht expression in  
 CC plants, conferring a dwarf phenotype on a plant which is correctable  
 CC by treatment with gibberellin. In addition, the products can be  
 CC used to produce Rht mutant plants which are dwarfed compared with  
 CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants  
 CC may be made by knocking out Rht or the relevant homologous gene in  
 CC the plant of interest. Plants may be made which are resistant to  
 CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,  
 CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds  
 CC dwarf but let crop plants grow tall. The present sequence is encoded by  
 CC the partial sequence of the maize D8-2023 allele.  
 CC  
 SQ Sequence 123 AA;  
 QY 1 LNAPPPPLPPA 11  
 DB 82 lnappaplp 92  
 Query Match 73.7%; Score 56; DB 20; Length 123;  
 Best Local Similarity 90.9%; Pred. No. 10;  
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 RESULT 11  
 ID AAY02541 standard; Protein; 630 AA.  
 AC AAY02541;  
 XX  
 DT 16-JUL-1999 (first entry)  
 DE Protein encoded by maize 1a1 genomic clone sequence.  
 XX  
 DE Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;  
 KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;  
 KW paclobutrazol; maize.  
 XX  
 OS Zea mays.  
 OS  
 PN WO9909174-A1.  
 XX  
 PD 25-FEB-1999.  
 XX  
 PE 07-AUG-1998; 98WO-GB02383.  
 XX  
 PR 13-AUG-1997; 97GB-0017192.  
 XX  
 PA (PLAN-) PLANT BIOSCIENCE LTD.  
 XX  
 PI Harberd NP, Peng J, Richards DE;  
 XX  
 DR WPI: 1999-181040/15.  
 DR N-PSDB; AAX36280.  
 XX  
 PT New Triticum aestivum polynucleotides - encode a polypeptide which  
 PT provides inhibition of the growth of plants, which inhibition is  
 PT antagonised by gibberellin, used to confer a dwarf phenotype  
 XX  
 PS Disclosure: Fig 9d; 88pp; English.  
 XX

CC The specification describes polypeptides encoded by the Rht gene (and  
 CC its homologues) that, when expressed in Triticum aestivum, inhibit  
 CC growth of the plant. This growth inhibition is antagonised by  
 CC gibberellin. The products can be used to provide Rht expression in  
 CC plants, conferring a dwarf phenotype on a plant which is correctable  
 CC by treatment with gibberellin. In addition, the products can be  
 CC used to produce Rht mutant plants which are dwarfed compared with  
 CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants  
 CC may be made by knocking out Rht or the relevant homologous gene in  
 CC the plant of interest. Plants may be made which are resistant to  
 CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,  
 CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds  
 CC dwarf but let crop plants grow tall. The present sequence is encoded by  
 CC the maize 1a1 genomic clone sequence.  
 CC  
 SQ Sequence 630 AA;  
 QY 1 LNAPPPPLPPA 11  
 DB 110 lnappaplp 120  
 Query Match 73.7%; Score 56; DB 20; Length 630;  
 Best Local Similarity 90.9%; Pred. No. 42;  
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 RESULT 12  
 ID AAB43674 standard; Protein; 293 AA.  
 AC AAB43674;  
 XX  
 DT 08-FEB-2001 (first entry)  
 DE Human cancer associated protein sequence SEQ ID NO:1119.  
 XX  
 DE Human; cancer associated gene; cancer antigen; detection; cancer;  
 KW diagnosis; cytostatic; proliferative; vulnerability; immunomodulator;  
 KW antidiabetic; antidiabetic; antirheumatic; antidiabetic; antiviral;  
 KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiac;  
 KW dermatological; neuroprotective; thrombolytic; coagulant; neotropic;  
 KW vasotropic; antiproliferative; angiogenic; gene therapy; inflammation;  
 KW immune disorder; haematopoietic cell disorder; autoimmune disorder;  
 KW allergic reaction; graft versus host disease; organ rejection;  
 KW haemostatic; thrombolytic; cardiovascular disorder; infection;  
 KW neurological disease; drug screening.  
 XX  
 OS Homo sapiens.  
 OS  
 PN WO200055350-A1.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PE 08-MAR-2000; 2000MO-US05882.  
 XX  
 PR 12-MAR-1999; 99US-0124270.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM;  
 XX  
 DR WPI: 2000-587533/55.  
 DR N-PSDB; AAC77883.  
 XX  
 PT Novel isolated nucleic acids comprising sequences encoding peptides  
 PT useful for treating or diagnosing e.g. cancer -  
 XX  
 PS Claim 11; Page 1730-1731; 2352pp; English.  
 XX  
 CC AAC77607 to AAC78448 encode the human cancer associated proteins given  
 CC in AAB43398 to AAB44239. The proteins can have activities based on the  
 CC tissues and cells the genes are expressed in. Example of activities

CC include: cytostatic; proliferative; vulnerary; immunomodulator;  
 CC antidiabetic; antiasthmatic; antirheumatic; antiarthritic;  
 CC antiinflammatory; antihypertoid; antiallergic; antibacterial; antiviral;  
 CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;  
 CC neotropic; vasotropic; antipsoriatic and antiangiogenic. The  
 CC polynucleotides and polypeptides can be used for preventing, treating or  
 CC ameliorating medical conditions and diagnosing pathological conditions.  
 CC Polynucleotides, antibodies, agonists and antagonists from  
 CC the present invention may be used to treat immune disorders by activating  
 CC or inhibiting the proliferation, differentiation or mobilisation of  
 CC immune cells, to treat disorders of haematopoietic cells, autoimmune  
 CC disorders, allergic reactions, graft versus host disease and organ  
 CC rejection, modulate haemostatic or thrombolytic activity, modulate  
 CC inflammation, cancers, cardiovascular disorders, neurological disease and  
 CC bacterial or viral infections. The peptides, nucleotides, antibodies,  
 CC agonists and antagonists may be also be used in drug screens. AAC78449 to  
 CC AAC78457 and AAB44240 represent sequences used in the exemplification of  
 CC the present invention.  
 CC  
 XX Sequence 293 AA:

Query Match 72.4%; Score 55; DB 21; Length 293;  
 Best Local Similarity 90.0%; Pred. No. 28;  
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 PPPPLPPAPQ 13  
 |||||  
 Db 16 PPPPPPPAPQ 25

RESULT 13  
 AAB79117  
 ID AAB79117 standard; Protein: 313 AA.

XX AAB79117;  
 DT 30-APR-2001 (first entry)

DE Corynebacterium glutamicum HA protein sequence SEQ ID NO:190.

KM Corynebacterium glutamicum; homeostasis; adaptation; HA protein;  
 KM fine chemical production; organic acid; proteinogenic amino acid;  
 KM nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;  
 KM nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;  
 KM carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;  
 KM diagnosis; Corynebacterium diptheriae; genetic engineering;  
 KM Brevibacterium; environmental condition.

OS Corynebacterium glutamicum.

PN WO200100842-A2.

PD 04-JAN-2001.

PF 23-JUN-2000; 2000WO-IB00911.

XX 25-JUN-1999; 99US-0141031.  
 PR 08-JUL-1999; 99DE-1031636.  
 PR 09-JUL-1999; 99DE-1032125.  
 PR 09-JUL-1999; 99DE-1032126.  
 PR 09-JUL-1999; 99DE-1032127.  
 PR 09-JUL-1999; 99DE-1032128.  
 PR 09-JUL-1999; 99DE-1032129.  
 PR 09-JUL-1999; 99DE-103226.  
 PR 14-JUL-1999; 99DE-1032920.  
 PR 14-JUL-1999; 99DE-1032922.  
 PR 14-JUL-1999; 99DE-1032924.  
 PR 14-JUL-1999; 99DE-1032928.  
 PR 14-JUL-1999; 99DE-1032930.  
 PR 14-JUL-1999; 99DE-1032933.  
 PR 14-JUL-1999; 99DE-1032935.  
 PR 14-JUL-1999; 99DE-1032973.

PR 14-JUL-1999; 99DE-1033002.  
 PR 14-JUL-1999; 99DE-1033003.  
 PR 14-JUL-1999; 99DE-1033005.  
 PR 14-JUL-1999; 99DE-1033006.  
 PR 31-AUG-1999; 99DE-1041378.  
 PR 31-AUG-1999; 99DE-1041379.  
 PR 31-AUG-1999; 99DE-1041390.  
 PR 31-AUG-1999; 99DE-1041391.  
 PR 03-SEP-1999; 99DE-1042088.

PA (BADI ) BASF AG.

XX Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;

XX WPI; 2001-061974/07.

DR N-PSDB; AAF71232.

PT New isolated Corynebacterium glutamicum nucleic acid for production or  
 PT modulation of production of fine chemicals such as amino acids,  
 PT nucleosides, nucleotides, lipids, fatty acids, carbohydrates, vitamins  
 PT or enzymes -

PS Claim 20; Page 399-400; 712pp; English.

CC AAF71138 to AAF71357 encode the Corynebacterium glutamicum homeostasis  
 CC and adaptation (HA) proteins given in AAB79023 to AAB79242. The  
 CC C. glutamicum HA genes (I) can be used in vectors for expression in host  
 CC cells and production of fine chemicals, such as, an organic acid,  
 CC proteinogenic or nonproteinogenic amino acid (preferred), purine or  
 CC pyrimidine base, nucleoside, nucleotide, lipid, saturated or unsaturated  
 CC fatty acid, diol, carbohydrate, aromatic compound, vitamin, cofactor,  
 CC polypeptide or enzyme. The amino acids produced can be lysine, glutamine,  
 CC glutamate, valine, leucine, isoleucine, arginine, proline, histidine,  
 CC cysteine, alanine, aspartate, glycine, serine, threonine, methionine,  
 CC tyrosine, phenylalanine, or tryptophan. The fine chemical production can  
 CC be modulated. The presence of (I) or HA proteins encoded by them are  
 CC used for diagnosing the presence or activity of Corynebacterium  
 CC diptheriae. (I) can be used to map the C. glutamicum genome or can be  
 CC used as markers for genetically engineered Corynebacterium or  
 CC Brevibacterium. The HA proteins encoded by the (I) are used to maintain  
 CC homeostasis in C. glutamicum or help the microorganism to adapt to  
 CC different environmental conditions.

SO Sequence 313 AA;

Query Match 72.4%; Score 55; DB 22; Length 313;  
 Best Local Similarity 90.0%; Pred. No. 30;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 LNAPPPPLPP 10  
 |:|||||  
 Db 288 l0pppplpp 297

RESULT 14  
 AAB79242  
 ID AAB79242 standard; Protein: 313 AA.

AC AAB79242;

DT 30-APR-2001 (first entry)

DE Corynebacterium glutamicum HA protein sequence SEQ ID NO:440.

KM Corynebacterium glutamicum; homeostasis; adaptation; HA protein;  
 KM fine chemical production; organic acid; proteinogenic amino acid;  
 KM nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;  
 KM nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;  
 KM carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;  
 KM diagnosis; Corynebacterium diptheriae; genetic engineering;  
 KM Brevibacterium; environmental condition.

XX	Corynebacterium glutamicum.
PN	W0200100842-A2.
XX	
PD	04-JAN-2001.
XX	
PF	23-JUN-2000; 2000MO-IBO0911.
PR	25-JUN-1999; 990S-0141031.
PR	08-JUL-1999; 99DE-1031636.
PR	09-JUL-1999; 99DE-1032125.
PR	09-JUL-1999; 99DE-1032126.
PR	09-JUL-1999; 99DE-1032127.
PR	09-JUL-1999; 99DE-1032128.
PR	09-JUL-1999; 99DE-1032129.
PR	09-JUL-1999; 99DE-1032226.
PR	14-JUL-1999; 99DE-1032920.
PR	14-JUL-1999; 99DE-1032922.
PR	14-JUL-1999; 99DE-1032924.
PR	14-JUL-1999; 99DE-1032928.
PR	14-JUL-1999; 99DE-1033003.
PR	14-JUL-1999; 99DE-1033005.
PR	14-JUL-1999; 99DE-1033006.
PR	31-AUG-1999; 99DE-1041378.
PR	31-AUG-1999; 99DE-1041379.
PR	31-AUG-1999; 99DE-1041380.
PR	31-AUG-1999; 99DE-1041391.
PR	03-SEP-1999; 99DE-1042088.
XX	
PA	(BADI ) BASF AG.
PI	Pompejus M, Kroegeer B, Schroeder H, Zelder O, Habermann G;
XX	
DR	WFI; 2001-061974/07.
DR	N-PDB; AAF71357.
XX	
PT	New isolated Corynebacterium glutamicum nucleic acid for production or
PT	modulation of production of fine chemicals such as amino acids,
PT	nucleosides, nucleotides, lipids, fatty acids, carbohydrates, vitamins
XX	or enzymes -
PS	Claim 20: Page 711-712; 712pp; English.
CC	AAE71138 to AAE71357 encode the Corynebacterium glutamicum homeostasis
CC	and adaptation (HA) proteins given in AAB79023 to AAB79242. The
CC	C. glutamicum HA genes (I) can be used in vectors for expression in host
CC	cells and production of fine chemicals, such as, an organic acid,
CC	proteogenic or nonproteogenic amino acid (preferred), purine or
CC	pyrimidine base, nucleoside, nucleotide, lipid, saturated or unsaturated
CC	fatty acid, diol, carbohydrate, aromatic compound, vitamin, cofactor,
CC	polyketide or enzyme. The amino acids produced can be lysine, glutamine,
CC	glutamate, alanine, aspartate, glycine, serine, threonine, methionine,
CC	cysteine, valine, leucine, isoleucine, arginine, proline, histidine,
CC	tyrosine, phenylalanine, or tryptophan. The fine chemical production can
CC	be modulated. The presence of (I) or HA proteins encoded by them are
CC	used for diagnosing the presence or activity of Corynebacterium
CC	diphtheriae. (I) can be used to map the C. glutamicum genome or can be
CC	used as markers for genetically engineered Corynebacterium or
CC	Brevibacterium. The HA proteins encoded by the (I) are used to maintain
CC	homeostasis in C. glutamicum or help the microorganism to adapt to
CC	different environmental conditions.
SQ	Sequence 313 AA;
XX	

Query Match 72.4%; Score 55; DB 22; Length 313;  
 Best Local Similarity 90.0%; Pred. NO. 30;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy	1	LNPPPLPP	10	
		:		
Db	288	ldappplpp	297	
RESULT 15				
ID	AA014025	AA014025 standard; Protein: 347 AA.		
AC	AA014025;			
AC				
DT	17-OCT-2000 (first entry)			
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 13734.			
KW	Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter termination sequence.			
XX	Arabidopsis thaliana.			
XX	EP1033405-A2.			
PD	06-SEP-2000.			
XX				
PE	25-FEB-2000; 2000EP-0301439.			
XX				
PR	25-FEB-1999;	99US-0121825.		
PR	05-MAR-1999;	99US-0123180.		
PR	09-MAR-1999;	99US-0123548.		
PR	23-MAR-1999;	99US-0125788.		
PR	25-MAR-1999;	99US-0126264.		
PR	29-MAR-1999;	99US-0126785.		
PR	01-APR-1999;	99US-0127462.		
PR	06-APR-1999;	99US-0128234.		
PR	16-APR-1999;	99US-0128714.		
PR	19-APR-1999;	99US-0130077.		
PR	21-APR-1999;	99US-0130449.		
PR	23-APR-1999;	99US-0130510.		
PR	23-APR-1999;	99US-0130891.		
PR	28-APR-1999;	99US-0131449.		
PR	30-APR-1999;	99US-0132048.		
PR	30-APR-1999;	99US-0132407.		
PR	04-MAY-1999;	99US-0132484.		
PR	05-MAY-1999;	99US-0132485.		
PR	06-MAY-1999;	99US-0132486.		
PR	06-MAY-1999;	99US-0132487.		
PR	07-MAY-1999;	99US-0132863.		
PR	11-MAY-1999;	99US-0134256.		
PR	14-MAY-1999;	99US-0134218.		
PR	14-MAY-1999;	99US-0134219.		
PR	14-MAY-1999;	99US-0134221.		
PR	14-MAY-1999;	99US-0134370.		
PR	18-MAY-1999;	99US-0134768.		
PR	19-MAY-1999;	99US-0134941.		
PR	20-MAY-1999;	99US-0135124.		
PR	21-MAY-1999;	99US-0135353.		
PR	24-MAY-1999;	99US-0135629.		
PR	25-MAY-1999;	99US-0136021.		
PR	27-MAY-1999;	99US-0136392.		
PR	28-MAY-1999;	99US-0136782.		
PR	01-JUN-1999;	99US-0137222.		
PR	03-JUN-1999;	99US-0137528.		
PR	04-JUN-1999;	99US-0137502.		
PR	07-JUN-1999;	99US-0137724.		
PR	08-JUN-1999;	99US-0138094.		
PR	10-JUN-1999;	99US-0138540.		
PR	10-JUN-1999;	99US-0138847.		
PR	14-JUN-1999;	99US-0139119.		
PR	16-JUN-1999;	99US-0139452.		
PR	16-JUN-1999;	99US-0139453.		

PR 17-JUN-1999;	99US-0139492.	PR 13-AUG-1999;	99US-0148684.
PR 18-JUN-1999;	99US-0139454.	PR 16-AUG-1999;	99US-0149368.
PR 18-JUN-1999;	99US-0139455.	PR 17-AUG-1999;	99US-0149175.
PR 18-JUN-1999;	99US-0139456.	PR 18-AUG-1999;	99US-0149426.
PR 18-JUN-1999;	99US-0139457.	PR 20-AUG-1999;	99US-0149722.
PR 18-JUN-1999;	99US-0139458.	PR 20-AUG-1999;	99US-0149723.
PR 18-JUN-1999;	99US-0139459.	PR 20-AUG-1999;	99US-0149929.
PR 18-JUN-1999;	99US-0139460.	PR 23-AUG-1999;	99US-0149929.
PR 18-JUN-1999;	99US-0139461.	PR 23-AUG-1999;	99US-0149930.
PR 18-JUN-1999;	99US-0139462.	PR 25-AUG-1999;	99US-0150566.
PR 18-JUN-1999;	99US-0139463.	PR 26-AUG-1999;	99US-0150884.
PR 18-JUN-1999;	99US-0139750.	PR 27-AUG-1999;	99US-0151065.
PR 18-JUN-1999;	99US-0139763.	PR 27-AUG-1999;	99US-0151066.
PR 21-JUN-1999;	99US-0139817.	PR 27-AUG-1999;	99US-0151080.
PR 22-JUN-1999;	99US-0139839.	PR 30-AUG-1999;	99US-0151303.
PR 23-JUN-1999;	99US-0140353.	PR 31-AUG-1999;	99US-0151438.
PR 23-JUN-1999;	99US-0140354.	PR 01-SEP-1999;	99US-0151930.
PR 24-JUN-1999;	99US-0140635.	PR 07-SEP-1999;	99US-0152363.
PR 28-JUN-1999;	99US-0140823.	PR 10-SEP-1999;	99US-0153070.
PR 29-JUN-1999;	99US-0140991.	PR 13-SEP-1999;	99US-0153758.
PR 30-JUN-1999;	99US-0141287.	PR 15-SEP-1999;	99US-0154018.
PR 01-JUL-1999;	99US-0141842.	PR 16-SEP-1999;	99US-0154039.
PR 01-JUL-1999;	99US-0142154.	PR 20-SEP-1999;	99US-0154779.
PR 02-JUL-1999;	99US-0142055.	PR 22-SEP-1999;	99US-0155139.
PR 06-JUL-1999;	99US-0142390.	PR 23-SEP-1999;	99US-0155486.
PR 08-JUL-1999;	99US-0142803.	PR 24-SEP-1999;	99US-0155659.
PR 09-JUL-1999;	99US-0142920.	PR 28-SEP-1999;	99US-0156458.
PR 12-JUL-1999;	99US-0142977.	PR 29-SEP-1999;	99US-0156596.
PR 13-JUL-1999;	99US-0143542.	PR 04-OCT-1999;	99US-0157117.
PR 14-JUL-1999;	99US-0143624.	PR 05-OCT-1999;	99US-0157753.
PR 15-JUL-1999;	99US-0144005.	PR 06-OCT-1999;	99US-0157865.
PR 16-JUL-1999;	99US-0144085.	PR 07-OCT-1999;	99US-0158029.
PR 16-JUL-1999;	99US-0144086.	PR 08-OCT-1999;	99US-0158232.
PR 19-JUL-1999;	99US-0144325.	PR 12-OCT-1999;	99US-0158369.
PR 19-JUL-1999;	99US-0144331.	PR 13-OCT-1999;	99US-0159293.
PR 19-JUL-1999;	99US-0144332.	PR 13-OCT-1999;	99US-0159294.
PR 19-JUL-1999;	99US-0144333.	PR 13-OCT-1999;	99US-0159295.
PR 19-JUL-1999;	99US-0144334.	PR 14-OCT-1999;	99US-0159329.
PR 19-JUL-1999;	99US-0144335.	PR 14-OCT-1999;	99US-0159330.
PR 20-JUL-1999;	99US-0144352.	PR 14-OCT-1999;	99US-0159331.
PR 20-JUL-1999;	99US-0144632.	PR 14-OCT-1999;	99US-0159637.
PR 20-JUL-1999;	99US-0144884.	PR 14-OCT-1999;	99US-0159638.
PR 21-JUL-1999;	99US-0144814.	PR 18-OCT-1999;	99US-0159584.
PR 21-JUL-1999;	99US-0145086.	PR 21-OCT-1999;	99US-0160741.
PR 21-JUL-1999;	99US-0145085.	PR 21-OCT-1999;	99US-0160767.
PR 22-JUL-1999;	99US-0145085.	PR 21-OCT-1999;	99US-0160768.
PR 22-JUL-1999;	99US-0145087.	PR 21-OCT-1999;	99US-0160770.
PR 22-JUL-1999;	99US-0145192.	PR 21-OCT-1999;	99US-0160814.
PR 23-JUL-1999;	99US-0145145.	PR 22-OCT-1999;	99US-0160815.
PR 23-JUL-1999;	99US-0145218.	PR 22-OCT-1999;	99US-0160980.
PR 23-JUL-1999;	99US-0145224.	PR 22-OCT-1999;	99US-0160981.
PR 26-JUL-1999;	99US-0145276.	PR 22-OCT-1999;	99US-0160989.
PR 27-JUL-1999;	99US-0145913.	PR 25-OCT-1999;	99US-0161405.
PR 27-JUL-1999;	99US-0145918.	PR 25-OCT-1999;	99US-0161405.
PR 27-JUL-1999;	99US-0145919.	PR 25-OCT-1999;	99US-0161406.
PR 28-JUL-1999;	99US-0145951.	PR 26-OCT-1999;	99US-0161359.
PR 02-AUG-1999;	99US-0146386.	PR 26-OCT-1999;	99US-0161360.
PR 02-AUG-1999;	99US-0146388.	PR 26-OCT-1999;	99US-0161361.
PR 02-AUG-1999;	99US-0146389.	PR 28-OCT-1999;	99US-0161920.
PR 03-AUG-1999;	99US-0146389.	PR 28-OCT-1999;	99US-0161922.
PR 04-AUG-1999;	99US-0147038.	PR 28-OCT-1999;	99US-0161993.
PR 04-AUG-1999;	99US-0147204.	PR 29-OCT-1999;	99US-0162142.
PR 05-AUG-1999;	99US-0147302.		
PR 05-AUG-1999;	99US-0147320.		
PR 05-AUG-1999;	99US-0147260.		
PR 06-AUG-1999;	99US-0147303.		
PR 06-AUG-1999;	99US-0147416.		
PR 09-AUG-1999;	99US-0147935.		
PR 10-AUG-1999;	99US-0148171.		
PR 11-AUG-1999;	99US-0148319.		
PR 12-AUG-1999;	99US-0148341.		
PR 13-AUG-1999;	99US-0148565.		

Query Match

Best Local Similarity 72.4% Score 55; DB 21; Length 347;

Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 NAPPPLPPAP 12	1111111111
Db 151 nrppppppap 161	



Search completed: January 22, 2002, 16:26:00  
Job time: 140 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2002, 16:24:46 ; Search time 26.73 seconds

(without alignments)  
37.047 Million cell updates/sec

Title: US-09-485-529-103

Perfect score: 76

Sequence: 1 LNAPPPLPPAPQ 13

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR:68:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	56	73.7	599	2	T10798	pherophorin-S - Vo
2	55	72.4	267	2	T15645	hypothetical prote
3	55	72.4	367	2	T139172	cyclin-dependent k
4	55	72.4	446	2	A34418	H-2 region II bind
5	55	72.4	448	2	D41727	retinoid X recepto
6	55	72.4	451	2	A41651	retinoic acid rece
7	55	72.4	520	2	T84718	RXR-beta1 isoform
8	54	71.1	138	2	C96734	hypothetical prote
9	54	71.1	998	2	T30930	hypothetical prote
10	54	71.1	2453	2	S60254	nuclear receptor c
11	54	71.1	3119	2	T49729	HD protein - mouse
12	53	69.7	257	2	S32101	PHL5A protein - c
13	53	69.7	373	2	A70856	probable lppz prot
14	53	69.7	731	2	T19721	hypothetical prote
15	52	68.4	238	2	T32889	hypothetical prote
16	52	68.4	281	2	T38707	Fas ligand - human
17	52	68.4	359	2	T13478	hypothetical prote
18	52	68.4	389	2	S27200	proline-rich prote
19	52	68.4	497	2	T15812	hypothetical prote
20	52	68.4	543	2	S25128	61k protein - Auto
21	52	68.4	561	2	G96752	unknown protein F2
22	52	68.4	596	2	T03908	hypothetical prote
23	52	68.4	699	2	A96529	hypothetical prote
24	52	68.4	798	2	T19864	hypothetical prote
25	52	68.4	1132	2	A35098	MHC class III hist
26	52	68.4	1151	2	T18535	high molecular mas
27	52	68.4	1264	2	A36858	G2R protein - vari
28	52	68.4	1308	2	T15280	hypothetical prote
29	52	68.4	1388	2	T00063	hypothetical prote

30	52	68.4	1460	1	EDBE1F	Immediate-early pr
31	52	68.4	1560	2	T00080	hypothetical prote
32	52	68.4	1733	1	B45344	probable nuclear a
33	52	68.4	1872	2	T24683	hypothetical prote
34	52	68.4	1896	2	B72175	D1SR protein - var
35	52	68.4	1897	2	T28621	hypothetical prote
36	52	68.4	1958	2	B45055	hypothetical prote
37	52	68.4	2123	2	F86348	hypothetical prote
38	51	67.1	281	2	T29150	hypothetical prote
39	51	67.1	294	2	A55477	survival motor neu
40	51	67.1	399	2	A49754	homeobox protein -
41	51	67.1	433	2	T07910	hydroxyproline-ric
42	51	67.1	541	2	T19304	hypothetical prote
43	51	67.1	564	2	T45866	hypothetical prote
44	51	67.1	574	2	T05964	probable low-affin
45	51	67.1	666	2	B70803	hypothetical prote

#### ALIGNMENTS

RESULT 1  
T10798  
pherophorin-S - Volvox carteri  
C:Species: Volvox carteri  
C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000  
C:Accession: T10798  
R:Godl, K.; Hallmann, A.; Wenzl, S.; Sumper, M.  
EMBO J. 16, 25-34, 1997  
A:Title: Differential targeting of closely related ECM-glycoproteins: The pherophori  
A:Reference number: Z17154, PMID:97162277  
A:Accession: T10798  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-599 <GOD>  
A:Cross-references: EMBL:Y07752; NID:q1655698; PIDN:CAA69032.1; PID:q1655699  
A:Experimental source: strain HK 10; sub-species Nagariensis  
A>Note: In contrast to the other pherophorins, pherophorin-S is targeted to the cell-  
C:Keywords: extracellular matrix; glycoprotein; pheromone

Query Match 73.7%; Score 56; DB 2; Length 599;  
Best Local Similarity 81.8%; Pred. No. 16;  
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 NAPPPLPPAP 12  
Db 216 NAPPPLPPSP 226

RESULT 2  
T15645  
hypothetical protein C26F1.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: T15645  
R:Geisel, C.; Stellyes, L.; Bradshaw, H.  
submitted to the EMBL Data Library, March 1996  
A:Description: The sequence of C. elegans cosmid C26F1.  
A:Reference number: Z18381  
A:Accession: T15645  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-267 <GEI>  
A:Cross-references: EMBL:U53148; NID:q1255375; PID:q1255384; PIDN:AA37073.1; GSPDB:G  
A:Experimental source: strain Bristol N2; clone C26F1  
C:Genetics:  
A:Gene: C26F1.1  
A:Map position: 5  
A:Introns: 54/3; 69/2; 85/2; 156/2; 171/3; 192/3

Query Match 72.4%; Score 55; DB 2; Length 267;

Best Local Similarity 90.0%; Pred. No. 9.5;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 4 PPPPLPPAPQ 13  
||| |||||  
Db 90 PPPPPPPAPQ 99

## RESULT 3

139172  
cyclin-dependent kinase 5 activator isoform p391 - human  
C:Species: Homo sapiens (man)  
C:Date: 01-Mar-1996 #sequence\_revision 01-Mar-1996 #text\_change 21-Jul-2000  
C:Accession: 139172  
R:Tang, D.; Yeung, J.; Lee, K.; Matsushita, M.; Matsui, H.; Tomizawa, K.; Hatase, O.; Waj  
J. Biol. Chem. 270, 26897-26903, 1995  
A:Title: An isoform of the neuronal cyclin-dependent kinase 5 (cdk5) activator.  
A:Reference number: 139172; MUID:96070784  
A:Accession: 139172  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-367 <RES>  
A:Cross-references: EMBL:U34051; NID:g1063622; PIDN:AAC50278.1; PID:g1063623  
C:Superfamily: tau-protein kinase II 23K chain

Query Match 72.4%; Score 55; DB 2; Length 367;  
Best Local Similarity 90.0%; Pred. No. 13;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 PPPPLPPAPQ 13  
||| |||||  
Db 155 PPPPPPPAPQ 164

## RESULT 4

A34418  
H-2 region II binding protein - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 15-Jun-1990 #sequence\_revision 09-Oct-1992 #text\_change 20-Sep-1999  
C:Accession: A34418  
R:Hamada, K.; Gleason, S.L.; Levy, B.Z.; Hirschfeld, S.; Appella, E.; Ozato, K.  
Proc. Natl. Acad. Sci. U.S.A. 86, 8289-8293, 1989  
A:Title: H-2RIIbP, a member of the nuclear hormone receptor superfamily that binds to b  
A:Reference number: A34418; MUID:90046782  
A:Accession: A34418  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-446 <HAM>  
A:Cross-references: GB:M26804; NID:g193740; PIDN:AAA37772.1; PID:g387184  
C:Superfamily: unassigned erba-related proteins; erba transforming protein homology  
C:Keywords: DNA binding; transcription regulation; zinc finger  
F:116-367/Domain: erba transforming protein homology <ERBA>

Query Match 72.4%; Score 55; DB 2; Length 446;  
Best Local Similarity 72.7%; Pred. No. 16;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 NAPPPLPPAP 12  
:|||||:|  
Db 31 SAPPPPPPPP 41

## RESULT 5

D41727  
retinoid X receptor beta - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 17-Mar-1999  
C:Accession: D41727  
R:Leid, M.; Kastner, P.; Lyons, R.; Nakshatri, H.; Saunders, M.; Zacharski, T.; Chen,  
Cell 68, 377-395, 1992  
A:Title: Purification, cloning, and RXR identity of the HeLa cell factor with which RXR

A:Reference number: A41727; MUID:92127595  
A:Accession: D41727  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-448 <LEI>  
A:Experimental source: liver  
A:Note: sequence extracted from NCBI backbone (NCBI:77547)  
C:Genetics:  
A:Start codon: CTG  
C:Superfamily: unassigned erba-related proteins; erba transforming protein homology  
C:Keywords: zinc finger  
F:118-364/Domain: erba transforming protein homology <ERBA>

Query Match 72.4%; Score 55; DB 2; Length 448;  
Best Local Similarity 72.7%; Pred. No. 16;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 NAPPPLPPAP 12  
:|||||:|  
Db 33 SAPPPPPPPP 43

## RESULT 6

A41651  
retinoic acid receptor coregulator - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 28-May-1992 #sequence\_revision 28-May-1992 #text\_change 20-Sep-1999  
C:Accession: A41651  
R:Yu, V.C.; Delserre, C.; Andersen, B.; Holloway, J.M.; Devary, O.V.; Naeae, A.M.; Ki  
Cell 67, 1251-1266, 1991  
A:Title: RXRbeta: a coregulator that enhances binding of retinoic acid, thyroid hormo  
A:Reference number: A41651; MUID:92103690  
A:Accession: A41651  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-451 <YUA>  
A:Cross-references: GB:M81766; NID:g206602; PIDN:AAA42025.1; PID:g206603  
C:Superfamily: unassigned erba-related proteins; erba transforming protein homology  
C:Keywords: zinc finger  
F:121-367/Domain: erba transforming protein homology <ERBA>

Query Match 72.4%; Score 55; DB 2; Length 451;  
Best Local Similarity 72.7%; Pred. No. 16;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 NAPPPLPPAP 12  
:|||||:|  
Db 36 SAPPPPPPPP 46

## RESULT 7

184718  
RXR-beta1 isoform - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 20-Aug-1999  
C:Accession: 184718  
R:Nagata, T.; Kanno, Y.; Ozato, K.; Taketo, M.  
Gene 142, 183-189, 1994  
A:Title: The mouse RXRb gene encoding RXR beta: genomic organization and two mRNA iso  
A:Reference number: 184718; MUID:94252565  
A:Accession: 184718  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-520 <RES>  
A:Cross-references: GB:D21831; NID:g497828; PIDN:BA04858.1; PID:g987669  
C:Genetics:  
A:Gene: RXRb  
A:introns: 76/1; 148/3; 201/1; 362/1; 406/2; 437/1; 472/2  
C:Superfamily: retinoic acid receptor alpha; erba transforming protein homology  
F:190-436/Domain: erba transforming protein homology <ERBA>

Thu Jan 24 09:56:12 2002

us-09-485-529-103.rpr

Page 3

```

Query Match      72.4%; Score 55; DB 2; Length 520;
Best Local Similarity 72.7%; Pred. No. 18;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 NAPPPLPPAP 12
    :|||||:
Db 105 SAPPMPMP 115

RESULT 8
hypothetical protein FL5H1.18 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse ear cross)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: C96734
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Maitl, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: C96734
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-138 <STO>
A:Cross-references: GB:AE005173; NID:g5902409; PIDN:AAD5511.1; GSPDB:GN00141
C:Genetics:
A:Gene: FL5H1.18
A:Map position: 1

Query Match      71.1%; Score 54; DB 2; Length 138;
Best Local Similarity 68.9%; Pred. No. 6.4;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPAP 12
    :|||||:
Db 71 PPPPLPPSP 79

RESULT 9
hypothetical protein D1007.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
R:Davidson, S.; Rohlfing, T.
submitted to the EMBL Data Library, August 1999
A:Description: The sequence of C. elegans cosmid D1007.
A:Reference number: Z20934
A:Accession: T30930
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-998 <DAV>
A:Cross-references: EMBL:AF003151; PIDN:AAB54225.1
C:Genetics:
A:Experimental source: strain Bristol N2; clone D1007
A:Map position: 1
A:Introns: 54/3; 141/1; 314/2; 532/1; 711/3; 756/3; 907/2; 939/1
A:Note: D1007.7

Query Match      71.1%; Score 54; DB 2; Length 998;
Best Local Similarity 68.8%; Pred. No. 45;
Matches 11; Conservative 0; Mismatches 1; Indels 4; Gaps 1;

OY 1 LNAPPPP---LPPAP 12

```

```

Db 714 LNAPPPGIPGPPAP 729
    :|||||:
RESULT 10
S60254
nuclear receptor co-repressor N-COR - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 08-Oct-1999
C:Accession: S60254
R:Hoerlein, A.J.; Naeer, A.M.; Heinzel, T.; Torchia, J.; Gloss, B.; Kurokawa, R.; Ry
Nature 377, 397-404, 1995
A:Title: Ligand-independent repression by the thyroid hormone receptor mediated by a
A:Reference number: S60254; MUID:96008539
A:Accession: S60254
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-2453 <HOB>
A:Cross-references: EMBL:U35312; NID:g1022717; PIDN:AAB17125.1; PID:g1022718

Query Match      71.1%; Score 54; DB 2; Length 2453;
Best Local Similarity 80.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 4 PPPPLPPAPQ 13
    :|||||:
Db 606 PPPPLPPPE 615

RESULT 11
I49729
HD protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 05-Nov-1999
C:Accession: I49729; I49730
R:Lin, B.; Nasir, J.; Macdonald, H.; Hutchinson, G.; Graham, R.; Rommens, J.M.; Hayde
Hum. Mol. Genet. 3, 85-92, 1994
A:Title: Sequence of the murine huntington disease gene: evidence for conservation an
A:Reference number: I49729; MUID:94214482
A:Accession: I49729
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-3119 <RES>
A:Cross-references: GB:U23312; NID:g438804; PIDN:AAA37799.1; PID:g438805
A:Accession: I49730
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1521, 2002-3119 <RE2>
A:Cross-references: GB:U23313; NID:g438806; PIDN:AAA37800.1; PID:g438807
C:Keywords: alternative splicing

Query Match      71.1%; Score 54; DB 2; Length 3119;
Best Local Similarity 81.8%; Pred. No. 1.4e+02;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 APPPLPPAPQ 13
    :|||||:
Db 29 APPPPPPPPQ 39

RESULT 12
S32101
PHLP5A protein - common timothy (fragment)
C:Species: Phleum pratense (common timothy)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 03-Mar-1995
C:Accession: S32101
R:Butt, A.; Becker, W.M.; Petersen, A.; Schramm, G.; Schlaak, M.
submitted to the EMBL Data Library, February 1993
A:Description: Phl pv (timothy grass) major allergen bears at least two B-cell epitop
A:Reference number: S32101
A:Accession: S32101

```

A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-257 <BUF>  
A:Cross-references: EMBL:X70942  
C:Superfamily: grass pollen allergen IX

Query Match 69.7%; Score 53; DB 2; Length 257;  
Best Local Similarity 81.8%; Pred. No. 15;  
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 3 APPPLPPAPQ 13  
Db 228 AKPPLPPPPQ 238

RESULT 13  
A70856  
Probable lppz protein - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C:Accession: A70856  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrett, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987  
A:Accession: A70856  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-373 <COI>  
A:Cross-references: GB:AL021287; GB:AL123456; NID:93261508; PIDN:CAA16091.1; PID:el23775  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: lppz

Query Match 69.7%; Score 53; DB 2; Length 373;  
Best Local Similarity 88.9%; Pred. No. 22;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 4 PPPPLPPAP 12  
Db 51 PPPPLPPVP 59

RESULT 14  
T19721  
hypothetical protein C34E11.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000  
C:Accession: T19721  
R:McMurray, A.  
submitted to the EMBL Data Library, November 1995  
A:Reference number: Z19169  
A:Accession: T19721  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-731 <WIL>  
A:Cross-references: EMBL:Z67754; PIDN:CAA91752.1; GSPDB:GN00028; CESP:C34E11.2  
A:Experimental source: clone C34E11  
C:Genetics:  
A:Gene: CESP:C34E11.2  
A:Map position: X  
A:Introns: 54/3; 149/3; 183/1; 224/2; 273/2; 330/3; 378/3; 416/1; 510/1; 578/2; 616/2  
C:Superfamily: Caenorhabditis elegans hypothetical protein C34E11.2

Query Match 69.7%; Score 53; DB 2; Length 731;  
Best Local Similarity 61.5%; Pred. No. 43;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 LNAPPPLPPAPQ 13  
Db 494 MSSPPPLPPRPPQ 506

RESULT 15  
T32889  
hypothetical protein C34B2.9 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T32889  
R:Graves, T.; Suterer, C.; Hawkins, M.; Wilson, R.  
submitted to the EMBL Data Library, January 1998  
A:Description: The sequence of C. elegans cosmid C34B2.  
A:Reference number: Z21241  
A:Accession: T32889  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-238 <GRA>  
A:Cross-references: EMBL:AF043693; PIDN:AAB97541.1; GSPDB:GN00019; CESP:C34B2.9  
A:Experimental source: strain Bristol N2; clone C34B2  
C:Genetics:  
A:Gene: CESP:C34B2.9  
A:Map position: 1  
A:Introns: 35/3; 91/2; 200/2

Query Match 68.4%; Score 52; DB 2; Length 238;  
Best Local Similarity 88.9%; Pred. No. 18;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 4 PPPPLPPAP 12  
Db 132 PPPPLPPLP 140

Search completed: January 22, 2002, 16:24:48  
Job time: 68 sec

Thu Jan 24 09:56:12 2002

us-09-485-529-103.rpr

---

Page 5





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2002, 16:30:28 ; Search time 17.09 Seconds  
(without alignments)  
27.890 Million cell updates/sec

Title: US-09-485-529-103  
Perfect score: 76  
Sequence: 1 LNAPPPLPPAPQ 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 3664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SWISSProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	57	75.0	282	1	ATF5_HUMAN
2	57	75.0	901	1	Y298_HUMAN
3	55	72.4	367	1	CD55_HUMAN
4	55	72.4	369	1	CD55_MOUSE
5	55	72.4	458	1	RARB_RAT
6	55	72.4	520	1	RARB_MOUSE
7	54	71.1	591	1	NMT_MOUSE
8	54	71.1	2440	1	NCRL_HUMAN
9	54	71.1	2453	1	NCRL_MOUSE
10	54	71.1	3119	1	HD_MOUSE
11	53	69.7	724	1	CCT1_MOUSE
12	53	69.7	726	1	CCT1_HUMAN
13	53	69.7	727	1	CCT1_HORSE
14	52	68.4	281	1	PAS1_HUMAN
15	52	68.4	288	1	SMN_MOUSE
16	52	68.4	389	1	NDEP_MOUSE
17	52	68.4	408	1	G3PT_MOUSE
18	52	68.4	543	1	VP61_NPVAC
19	52	68.4	736	1	DVL2_MOUSE
20	52	68.4	809	1	CNAD_HUMAN
21	52	68.4	1132	1	BAT3_HUMAN
22	52	68.4	1164	1	PHOS_HUMAN
23	52	68.4	1733	1	VNOA_PRRNA
24	51	67.1	223	1	TOMB_XANCP
25	51	67.1	294	1	SMN_HUMAN
26	51	67.1	311	1	CDX2_HUMAN
27	51	67.1	372	1	PEXC_ARATH
28	51	67.1	399	1	HXA_MOUSE
29	51	67.1	744	1	IEB3_HCVAA
30	51	67.1	808	1	Y066_NPVAC
31	51	67.1	825	1	RCAL1_YEAST
32	51	67.1	884	1	ANDR_EULFC
33	51	67.1	2911	1	FRN2_HUMAN

## ALIGNMENTS

RESULT	ID	ATF5_HUMAN	STANDARD:	PRT:	282 AA.
AC	Q9Y2D1	Q9Y2D1	Q9Y2D1		
DT	20-AUG-2001	(Rel. 40, Last sequence update)			
DT	20-AUG-2001	(Rel. 40, Last annotation update)			
DE	CYCLOC-AMP-DEPENDENT TRANSCRIPTION FACTOR ATF-5 (ACTIVATING TRANSCRIPTION FACTOR 5) (TRANSCRIPTION FACTOR ATF5).				
GN	ATF5 OR ATF5.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]	SEQUENCE FROM N.A.			
RP	MEDLINE=20558615; PubMed=11087824;				
RX	White J.H., McIlhinney R.A.J., Wise A., Ciruela F., Chan W.Y.,				
RA	Emson P.C., Billington A., Marshall F.H.;				
RT	"The GABAB receptor interacts directly with the related transcription factors CREB2 and ATF5."				
RT	Proc. Natl. Acad. Sci. U.S.A. 97:13967-13972(2000).				
RL	[2]	SEQUENCE FROM N.A.			
RP	Kohroki J., Tanaka K.;				
RL	Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.				
RN	[3]	SEQUENCE OF 161-282 FROM N.A.			
RP	MEDLINE=9303793; PubMed=10373550;				
RX	Patl D., Weistrich M.L., Pion S.E.;				
RT	"Human Cdc34 and Rad6 ubiquitin-conjugating enzymes target repressors of cyclic AMP-induced transcription for proteolysis."				
RT	Mol. Cell. Biol. 19:3001-3013(1999).				
CC	-1- FUNCTION: THIS PROTEIN BINDS THE CAMP RESPONSE ELEMENT (CRE) (CONSENSUS: 5'/GTGACGT(A/C)(A/G)-3'), A SEQUENCE PRESENT IN MANY VIRAL AND CELLULAR PROMOTERS.				
CC	-1- SUBUNIT: BINDS DNA AS A DIMER (POTENTIAL). INTERACTS WITH GABAB RECEPTOR.				
CC	-1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).				
CC	-1- SIMILARITY: TO OTHER BZIP PROTEINS.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	EMBL: AF305687; AAG2558.1; -				
DR	EMBL: AB021663; BA878477.2; -				
DR	EMBL: AF101388; AAD28370.1; -				
DR	InterPro: IPR002637; Ham1p-like.				
DR	InterPro: IPR002965; P-f1ch_extensn.				
DR	InterPro: IPR001871; bzip.				
DR	Pfam: PF00170; bzip, 1.				

34	50	65.8	165	1	Y084_RICPR	Q9263 rickettsia
35	50	65.8	288	1	FXE3_MOUSE	Q99Y14 mus musculu
36	50	65.8	307	1	RB33_CAPEL	Q20365 caenorhabdi
37	50	65.8	309	1	HXA4_CHICK	P17277 gallus galli
38	50	65.8	401	1	HME1_MOUSE	P09065 mus musculu
39	50	65.8	415	1	ACRO_PIG	P08001 sus scrofa
40	50	65.8	440	1	G3PT_MOUSE	Q64467 mus musculu
41	50	65.8	487	1	EBN2_EBV	P12978 Epstein-bar
42	50	65.8	533	1	RARB_HUMAN	P28702 homo sapien
43	50	65.8	559	1	WAS1_HUMAN	Q92558 homo sapien
44	50	65.8	601	1	3BP1_MOUSE	P55194 mus musculu
45	50	65.8	736	1	DVL2_HUMAN	O14641 homo sapien

```
DR PRINTS: PR01217; PRICHEXTENSN.
DR Produm: PD004952; Hampr_1like; 1.
DR SMART: SMO0338; BRIZ; 1.
DR PROSITE: PS00036; BRIP_BASIC; 1.
KW Transcription regulation; DNA-binding; Activator; Nuclear protein;
KW Multigene family.
FT DOMAIN 123 139 POLY-PRO.
FT DOMAIN 186 194 POLY-PRO.
FT DNA_BIND 210 230 BASIC MOTIF.
FT DOMAIN 236 250 LEUCINE-ZIPPER (PROBABLE).
FT CONFLICT 161 163 LLA -> RHE (IN REF. 3).
SQ SEQUENCE 282 AA; 30674 MW; DDB2P907CA0215A0.CRC64;

Query Match
Best Local Similarity 100.0%; Score 57; DB 1; Length 282;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLPPAP 12
DB 131 PPPPLPPAP 139

RESULT 2
Y298_HUMAN STANDARD; PRT; 901 AA.
ID Y298_HUMAN
AC 015016;
DT 15-JUL-1998 (Rel. 36, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN KIAA0298 (FRAGMENT).
GN KIAA0298.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Nagase T., Ishikawa K.-I., Seki N., Nakajima D., Ohira M.,
RA Miyajima N., Kotani H., Nomura N., Ohara O.;
RA Miyajima N., Kotani H., Nomura N., Ohara O.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE OF 305-901 FROM N.A.
RC TISSUE=Brain;
RA MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AB002296; BAA20758.2; -
DR PROSITE: PS50119; ZF_BBOX; 2.
KW Hypothetical protein; Zinc-finger; Repeat.
FT NON_TER 1
FT ZN_FING 33 79 B BOX-TYPE 1.
FT ZN_FING 93 134 B BOX-TYPE 2.
FT DOMAIN 486 532 PRO-RICH.
SQ SEQUENCE 901 AA; 99510 MW; 8A5C91B04F5F556 CRC64;

Query Match
Best Local Similarity 75.0%; Score 57; DB 1; Length 901;
Matches 90.0%; Pred. No. 15;
```

```
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 PPPPLPPAPQ 13
DB 500 PPPPLPPAPQ 509

RESULT 3
CD55_HUMAN STANDARD; PRT; 367 AA.
ID CD55_HUMAN
AC Q13319;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CYCLIN-DEPENDENT KINASE 5 ACTIVATOR 2 PRECURSOR (CDK5 ACTIVATOR 2)
DE (CYCLIN-DEPENDENT KINASE 5 REGULATORY SUBUNIT 2) (P39) (P39I).
GN CDK5R2 OR NCK5A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Hippocampus;
RC MEDLINE=96070784; PubMed=7592934;
RA Tang D., Yeung J., Lee K.-Y., Matsushita M., Matsui H., Tomizawa K.,
RA Hatase O., Wang J.H.;
RT "An isoform of the neuronal cyclin-dependent kinase 5 (cdk5)
RT activator.";
RL J. Biol. Chem. 270:26897-26903(1995).
CC -1- FUNCTION: ACTIVATOR OF CDK5/TPKII.
CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC SUBUNIT AND A REGULATORY
CC SUBUNIT.
CC -1- TISSUE SPECIFICITY: BRAIN AND NEURON SPECIFIC.
CC -1- SIMILARITY: BELONGS TO THE CYCLIN-DEPENDENT KINASE 5 ACTIVATOR
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U34051; AAC50278.1; -
DR MIM: 603764; -
FT PROPEP 1 ? POTENTIAL.
FT CHAIN ? 367 CYCLIN-DEPENDENT KINASE 5 ACTIVATOR 2.
FT DOMAIN 75 78 POLY-LYS.
FT DOMAIN 130 133 POLY-ALA.
FT DOMAIN 155 163 POLY-PRO.
SQ SEQUENCE 367 AA; 38704 MW; D8CB7C8E0A8200 CRC64;

Query Match
Best Local Similarity 72.4%; Score 55; DB 1; Length 367;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 PPPPLPPAPQ 13
DB 155 PPPPLPPAPQ 164

RESULT 4
CD55_MOUSE STANDARD; PRT; 369 AA.
ID CD55_MOUSE
AC O35926; O35277;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE CYCLIN-DEPENDENT KINASE 5 ACTIVATOR 2 PRECURSOR (CDK5 ACTIVATOR 2)
DE (CYCLIN-DEPENDENT KINASE 5 REGULATORY SUBUNIT 2) (P39) (P39I).
```

```

GN CDR5R2 OR NCR5A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RX MEDLINE=98322126; PubMed=9655938;
RA Nilsen F., Baekstrom A., Bark C.;
RT "Molecular cloning and characterisation of a mouse gene encoding an
RT isoform of the neuronal cyclin-dependent kinase 5 (CDK5) activator.";
RL Biochim. Biophys. Acta 1398:371-376(1998).
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ;
RA Zheng M., Lem R.K.H.;
RT "Comparative analysis of gene expression of the cyclin-dependent
RT kinase 5 (cdk5) activators p35 and p39 in the rat CNS and embryonic
RT development revealed distinctive overlap with cdk5.";
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ACTIVATOR OF CDK5/TIPK1.
CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC SUBUNIT AND A REGULATORY
CC SUBUNIT.
CC -1- SIMILARITY: BELONGS TO THE CYCLIN-DEPENDENT KINASE 5 ACTIVATOR
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U90267; AAC53595.1; -
DR EMBL: AF016393; AAB69709.1; -
DR MGP: MGI:130828; Cdk5r2.
DR InterPro: IPR002965; P-rich_extensn.
DR PRINTS: PR01217; PRICHEXTENS.
DR PROPEP 1 ?
FT CHAIN ? 369 POTENTIAL.
FT DOMAIN ? 78 CYCLIN-DEPENDENT KINASE 5 ACTIVATOR 2.
FT DOMAIN 155 163 POLY-LYS.
FT DOMAIN 348 351 POLY-PRO.
FT CONFLICT 91 91 G->V (IN REF. 2).
FT CONFLICT 172 172 S->G (IN REF. 2).
FT CONFLICT 191 191 R->G (IN REF. 2).
FT CONFLICT 197 197 S->R (IN REF. 2).
FT CONFLICT 220 220 L->W (IN REF. 2).
FT CONFLICT 248 248 S->R (IN REF. 2).
FT CONFLICT 253 253 A->G (IN REF. 2).
SQ SEQUENCE 369 AA; 38914 MW; DA4EB1FD4D48CE66 CRC64;

Query Match 72.4%; Score 55; DB 1; Length 369;
Best Local Similarity 90.0%; Pred. No. 11;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 4 PPPPLPAPQ 13
DB 155 PPPPPAPQ 164
RESULT 5
RXRB_RAT STANDARD; PRT; 458 AA.
AC P49743;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RETINOIC ACID RECEPTOR RXR-BETA (NUCLEAR RECEPTOR CO-REGULATOR 1)
DE (FRAGMENT).
GN RXRB OR NR2B2 OR RCOR-1.

```

```

OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RX MEDLINE=92103690; PubMed=1662118;
RA Yu V.C., Delser C., Andersen B., Holloway J.M., Devary O.V.,
RA Naess A.M., Kim S.Y., Boutin J.M., Glass C.K., Rosenfeld M.G.;
RT "RXR beta: a coregulator that enhances binding of retinoic acid,
RT thyroid hormone, and vitamin D receptors to their cognate response
RT elements.";
RL Cell 67:1251-1266(1991).
CC -1- FUNCTION: INVOLVED IN RETINOIC ACID RESPONSE PATHWAY. BINDS
CC 9-CIS RETINOIC ACID (9C-RA).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC NR2 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M81766; AAA42025.1; ALT_INIT.
DR HSSP: P19793; 2NLT.
DR InterPro: IPR000536; Hormone_rec_1lg.
DR InterPro: IPR001628; zf-C4.
DR Pfam: PF00104; hormone_rec_1.
DR Pfam: PF00105; zf-C4; 1.
DR SMART: SM00430; HOL1.1.
DR SMART: SM00398; zNF_C4; 1.
DR PROSITE: PS00031; NUCLEAR_RECEPTOR_1.
DR KX zinc_finger; Multigene family.
KW zinc_finger; Multigene family.
FT FT 1 1
FT DOMAIN <1 129 MODULATING (BY SIMILARITY).
FT DNA_BIND 130 195 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 130 150 C4-TYPE.
FT ZN_FING 166 190 C4-TYPE.
FT DOMAIN 196 255 HINGE.
FT DOMAIN 256 458 LIGAND-BINDING (BY SIMILARITY).
SQ SEQUENCE 458 AA; 49078 MW; 22D11956B11DB4BE CRC64;

Query Match 72.4%; Score 55; DB 1; Length 458;
Best Local Similarity 72.7%; Pred. No. 13;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
OY 2 NAPPPLPAP 12
DB 43 SAPPMPPPP 53
RESULT 6
RXRB_MOUSE STANDARD; PRT; 520 AA.
AC P28704; P33243;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RETINOIC ACID RECEPTOR RXR-BETA (MHC CLASS I REGULATORY ELEMENT
DE BINDING PROTEIN H-2RIIBP).
GN RXRB OR NR2B2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

OX NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RC STRAIN-BALB/C; TISSUE=LIVER;  
 RX MEDLINE=94252565; PubMed=8194750;  
 RA Nagata T., Kanno Y., Ozato K., Taketo M.;  
 RT "The mouse Rarb gene encoding RXR beta: genomic organization and two  
 RT mRNA isoforms generated by alternative splicing of transcripts  
 RT initiated from CPG island promoters.";  
 RL Gene 142:183-189(1994).  
 RN [2]  
 RP SEQUENCE OF 73-520 FROM N.A. (LONG ISOFORM).  
 RC TISSUE=Embryo;  
 RX MEDLINE=92127595; PubMed=1310259;  
 RA Leid M., Kastner P., Lyons R., Nakshatri H., Saunders M.,  
 RA Zacharewski T., Chen J.Y., Staub A., Garnier J.M., Mader S.,  
 RA Chambon P.;  
 RT "Purification, cloning, and RXR identity of the HeLa cell factor with  
 RT which RAR or TR heterodimerizes to bind target sequences  
 RT efficiently.";  
 RL Cell 68:377-395(1992).  
 RN [3]  
 RP SEQUENCE OF 75-520 FROM N.A. (LONG ISOFORM).  
 RC TISSUE=Liver;  
 RX MEDLINE=90046782; PubMed=2554307;  
 RA Hamada K., Gleason S.L., Levi B.-Z., Hirschfeld S., Appella E.,  
 RA Ozato K.;  
 RT "H-2RIIB, a member of the nuclear hormone receptor superfamily that  
 RT binds to both the regulatory element of major histocompatibility  
 RT class I genes and the estrogen response element.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:8289-8293(1989).  
 RN [4]  
 RP SEQUENCE FROM N.A. (SHORT ISOFORM).  
 RC TISSUE=Liver;  
 RX MEDLINE=92192447; PubMed=1312497;  
 RA Mangelsdorf D.J., Borgmeyer U., Heyman R.A., Zhou J.Y., Ong E.S.,  
 RA Oro A.E., Kakizuka A., Evans R.M.;  
 RT "Characterization of three RXR genes that mediate the action of 9-cis  
 RT retinoic acid.";  
 RL Genes Dev. 6:329-344(1992).  
 CC -1- FUNCTION: INVOLVED IN RETINOIC ACID RESPONSE PATHWAY. BINDS 9-CIS  
 CC RETINOIC ACID (9C-RA). BINDS TO BOTH, CLASS I REGULATORY ELEMENT  
 CC OF MAJOR HISTOCOMPATIBILITY CLASS I GENES (CIR) AND THE ESTROGEN  
 CC RESPONSE ELEMENT (ERE).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN  
 CC HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: IN ALL TISSUES TESTED INCLUDING BRAIN, THYMUS  
 CC SPLEEN AND LIVER.  
 CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,  
 CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.  
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.  
 CC NM2 SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: D21831; BAA04858.1; -;  
 DR EMBL: D21830; BAA04858.1; JOINED.  
 DR EMBL: D21831; BAA04859.1; -;  
 DR EMBL: D21830; BAA04859.1; JOINED.  
 DR EMBL: M84818; AAA40081.2; -;  
 DR EMBL: M26804; AAA37772.1; -;  
 DR EMBL: X66224; CAA46963.1; -;  
 DR PIR: A34418; A34418.  
 DR HSSP: P19793; 2NCL.  
 DR MGD: MGI:98215; Rxrb.  
 DR InterPro: IPR000536; Hormone\_rec\_1lg.

DR InterPro: IPR001723; Sterdhormone\_receptor.  
 DR InterPro: IPR001628; zf-C4.  
 DR Pfam: PF00104; hormone\_rec. 1.  
 DR Pfam: PF00105; zf-C4; 1.  
 DR PRINTS: PR00047; STROIDFINGER.  
 DR PRINTS: PR00398; STROHORMONER.  
 DR PRINTS: PR00545; RETINOIDXR.  
 DR SMART: SM00430; HOL1; 1.  
 DR SMART: SM00399; znf\_C4; 1.  
 DR PROSITE: PS00031; NUCLEAR\_RECEPTOR; 1.  
 DR Receptor: Transcription regulation; DNA-binding; Nuclear protein;  
 KW Zinc-finger; Multigene family; Alternative splicing.  
 FT DOMAIN 1 191  
 FT DNA\_BIND 192 257  
 FT ZN\_FING 192 212  
 FT ZN\_FING 228 252  
 FT DOMAIN 258 317  
 FT DOMAIN 318 520  
 FT VARSPLIC 1 110  
 FT CONFLICT 73 75  
 FT CONFLICT 75 77  
 FT CONFLICT 433 433  
 FT SEQUENCE 520 AA: 55865 MW: 71FAFAECDD2B505 CXC64;  
 SO  
 Query Match 72.4%; Score 55; DB 1; Length 520;  
 Best Local Similarity 72.7%; Pred. No. 15;  
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 NAPPPLPPAP 12  
 DB 105 SAPPPMPPPP 115  
 :|||||:|  
 RESULT 7  
 MNT\_MOUSE STANDARD; PRT; 591 AA.  
 ID MNT\_MOUSE  
 AC 008789; P97349; Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE MAX BINDING PROTEIN MNT (ROX PROTEIN) (MYC ANTAGONIST MNT).  
 GN MNT OR ROX.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RX MEDLINE=97152466; PubMed=9000049;  
 RA Hurlin P.J., Queva C., Eisenman R.N.;  
 RT "Mnt, a novel Max-interacting protein is coexpressed with Myc in  
 RT proliferating cells and mediates repression at Myc binding sites.";  
 RL Genes Dev. 11:44-58(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RX MEDLINE=97327566; PubMed=9184233;  
 RA Teroni G., Raymond A., Alcalay M., Borsani G., Tanigami A.,  
 RA Tonlorenzi R., Lo Nigro C., Messali S., Zollo M., Ledbetter D.H.,  
 RA Brent R., Ballabio A., Carrozzo R.;  
 RT "Rox, a novel bHLHZip protein expressed in quiescent cells that  
 RT heterodimerizes with Max, binds a non-canonical E box and acts as a  
 RT transcriptional repressor.";  
 RL EMO J. 16:2892-2906(1997).  
 CC -1- FUNCTION: BINDS DNA AS A HETERODIMER WITH MAX AND REPRESSSES  
 CC TRANSCRIPTION. BINDS TO THE CANONICAL E BOX SEQUENCE 5'-CACGTG-3'  
 CC AND, WITH HIGHER AFFINITY, TO 5'-CAGCG-3'.  
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN. BINDS DNA AS AN HOMODIMER OR A HETERODIMER WITH MAX.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF

TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

CC EMBL: U77356; AAB8687.1; -;  
 CC EMBL: Y07609; CAA68878.1; -;  
 CC MGDI: MG1:109150; Mat.  
 CC InterPro: IPR003015; HLH\_MYC.  
 CC InterPro: IPR001092; HLH\_dlm.  
 CC Pfam: PF00010; HLH; 1.  
 CC SMART: SM00353; HLH; 1.  
 CC PROSITE: PS00038; HELIX\_LOOP\_HELIX; FALSE\_NEG.  
 CC Transcription regulation; Repressor; Nuclear protein; DNA-binding.  
 CC DNA\_BIND 224 235 BASIC MOTIF (POTENTIAL).  
 CC DOMAIN 236 272 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
 CC DOMAIN 273 301 LEUCINE-ZIPPER (POTENTIAL).  
 CC CONFLICT 379 379 T -> P (IN REF. 2).  
 CC CONFLICT 392 392 V -> A (IN REF. 2).  
 CC CONFLICT 402 403 EE -> OO (IN REF. 2).  
 CC CONFLICT 414 414 G -> A (IN REF. 2).  
 CC CONFLICT 431 431 V -> A (IN REF. 2).  
 CC CONFLICT 465 465 A -> P (IN REF. 2).  
 CC CONFLICT 525 525 T -> A (IN REF. 2).  
 CC CONFLICT 558 558 G -> A (IN REF. 2).  
 CC SEQUENCE 591 AA; 63311 MW; 3FAE61A10CF4142F CRC64;

Query Match 71.1%; Score 54; DB 1; Length 591;  
 Best Local Similarity 90.0%; Pred. No. 21;  
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 PPPPLPPAPQ 13  
 DB 105 PPPPLPPAPQ 114

RESULT 8  
 NCRL\_HUMAN STANDARD; PRT; 2440 AA.  
 AC 075376; Q90UPV5; Q90Q18;  
 DT 20-AUG-2001 (Rel. 40; Created)  
 DT 20-AUG-2001 (Rel. 40; Last sequence update)  
 DT 20-AUG-2001 (Rel. 40; Last annotation update)  
 DE NUCLEAR RECEPTOR CO-REPRESSOR 1 (N-COR1) (N-COR).  
 GN NCOR1 OR KIAA1047.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Fetal brain;  
 RX MEDLINE=96393736; PubMed=9724795;  
 RA Wang J., Hoshino T., Redner R.L., Kajigaya S., Liu J.M.;  
 RT "ETO, fusion partner in t(8;21) acute myeloid leukemia, represses  
 RT transcription by interaction with the human N-COR/msin3/HDAC1  
 RT complex.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:10860-10865(1998).  
 RN [2]  
 RP SEQUENCE OF 782-2440 FROM N.A.  
 RC TISSUE-brain;  
 RX MEDLINE=99397452; PubMed=10470851;  
 RA Kikuno R., Nagase T., Ishikawa K.-I., Hirosewa M., Miyajima N.,  
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XIV.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro.";

RL DNA Res. 6:197-205(1999).  
 RN [3]  
 RP SEQUENCE OF 974-2440 FROM N.A.  
 RX MEDLINE=99375328; PubMed=10444336;  
 RA Nagaya T., Chen K.-S., Fujieda M., Ohmori S., Richer J.K.,  
 RA Horvitz K.B., Lupski J.R., Seo H.;  
 RT "Localization of the human nuclear receptor co-repressor (hn-COR) gene  
 RT between the CMT1a and the SMS critical regions of chromosome  
 RT 17p11.2.";  
 RL Genomics 59:339-341(1999).  
 CC -1- FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME  
 CC NUCLEAR RECEPTORS BY PROMOTING CHROMATIN CONDENSATION, THUS  
 CC PREVENTING ACCESS OF THE BASAL TRANSCRIPTION.  
 CC -1- SUBUNIT: FORMS A LARGE COREPRESSOR COMPLEX THAT CONTAINS SIN3A/B  
 CC AND HISTONE DEACETYLASES HDAC1 AND HDAC2. THIS COMPLEX ASSOCIATES  
 CC WITH THE THYROID (TR) AND THE RETINOIC ACID RECEPTORS (RAR) IN THE  
 CC ABSENCE OF LIGAND.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 CC -1- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT  
 CC ARE DIVIDED INTO THREE INDEPENDANT REPRESSION DOMAINS (RD1, RD2  
 CC AND RD3). THE C-TERMINAL REGION CONTAINS THE NUCLEAR RECEPTOR-  
 CC INTERACTING DOMAINS THAT ARE DIVIDED IN TWO SEPARATE INTERACTION  
 CC DOMAINS (ID1 AND ID2).  
 CC -1- DOMAIN: THE TWO INTERACTION DOMAINS (ID) CONTAIN A CONSERVED  
 CC SEQUENCE REFERRED TO AS THE CORNR BOX. THIS MOTIF IS REQUIRED AND  
 CC SUFFICIENT TO PERMIT BINDING TO UNLIGANDED TR AND RAR. SEQUENCES  
 CC FLANKING THE CORNR BOX DETERMINE NUCLEAR HORMONE RECEPTOR  
 CC SPECIFICITY.  
 CC -1- SIMILARITY: CONTAINS 1 SANT DOMAIN (COMPOSED OF SANT-A+SANT-B).  
 CC -1- SIMILARITY: CONTAINS 2 CORNR BOX.  
 CC -1- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS  
 CC FAMILY.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

CC EMBL: AF044209; AAC33550.1; -;  
 CC EMBL: AB028970; BAA82999.1; -;  
 CC EMBL: AB019524; BAA75814.1; -;  
 CC MIM: 600849; -;  
 CC InterPro: IPR001005; MYB\_DNA\_bind.  
 CC Pfam: PF00249; MYB\_DNA-binding; 2.  
 CC SMART: SM00395; SANT; 2.  
 CC PROSITE: PS50090; MYB\_3; 1.  
 CC Nuclear protein; Transcription regulation; DNA-binding; Repressor;  
 CC coiled coil.  
 CC FT 174 216 COILED COIL (POTENTIAL).  
 CC FT- DOMAIN 254 312 INTERACTION WITH SIN3A/B.  
 CC FT- DOMAIN 299 328 COILED COIL (POTENTIAL).  
 CC FT- DOMAIN 437 482 SANT-A (POTENTIAL).  
 CC FT- DNA\_BIND 625 670 SANT-B (POTENTIAL).  
 CC FT- DOMAIN 501 557 COILED COIL (POTENTIAL).  
 CC FT- DOMAIN 607 617 PRO-RICH.  
 CC FT- DOMAIN 988 1816 INTERACTION WITH ETO.  
 CC FT- DOMAIN 2055 2059 CORNR BOX OF ID1.  
 CC FT- DOMAIN 2263 2267 CORNR BOX OF ID2.  
 CC FT- DOMAIN 58 64 POLY-GLN.  
 CC FT- DOMAIN 593 603 POLY-ALA.  
 CC FT- DOMAIN 1032 1035 POLY-PRO.  
 CC FT- DOMAIN 1707 1712 POLY-ALA.  
 CC FT- DOMAIN 1952 1963 POLY-SER.  
 CC FT- CONFLICT 1014 1014 L -> V (IN REF. 2).  
 CC FT- CONFLICT 1508 1509 PP -> SS (IN REF. 2).  
 CC FT- CONFLICT 1561 1561 W -> R (IN REF. 2).  
 CC FT- CONFLICT 1567 1567 Q -> H (IN REF. 2).  
 CC SEQUENCE 2440 AA; 270263 MW; 60A4D7964D00EDAB CRC64;

```

Query Match      71.1%; Score 54; DB 1; Length 2440;
Best Local Similarity 80.0%; Pred. No. 82;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      4 PPPPPAPQ 13
        ||||| 1:
Db      607 PPPPPPEP 616

RESULT 9
NCRL_MOUSE
ID NCRL_MOUSE STANDARD: PRT: 2453 AA.
AC Q60974; Q60812;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NUCLEAR RECEPTOR CO-REPRESSOR 1 (N-COR1) (N-COR) (RETINOID X RECEPTOR
DE INTERACTING PROTEIN 13) (RIP13).
OS NCOR1 OR RXRIP13.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
RC TISSUE-Pituitary;
RX MEDLINE=96008539; PubMed=7566114;
RA Hoerlein A.J., Naeae A.M., Heinzel T., Torchia J., Gloss B.,
RA Kurokawa R., Ryan A., Kamel Y., Soederstrom M., Glass C.K.,
RA Rosenfeld M.G.;
RT "Ligand-independent repression by the thyroid hormone receptor
RT mediated by a nuclear receptor co-repressor.";
RL Nature 377:397-404(1995).
RN [2]
RP SEQUENCE OF 1792-2453 FROM N.A. (ISOFORM LONG).
RC TISSUE-Liver;
RX MEDLINE=95280959; PubMed=7760852;
RA Seol W., Choi H.S., Moore D.D.;
RT "Isolation of proteins that interact specifically with the retinoid X
RT receptor: two novel orphan receptors.";
RL Mol. Endocrinol. 9:72-85(1995).
CC -1- FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME
CC NUCLEAR RECEPTORS BY PROMOTING CHROMATIN CONDENSATION, THUS
CC PREVENTING ACCESS OF THE BASAL TRANSCRIPTION.
CC -1- SUBUNIT: FORMS A LARGE COREPRESSOR COMPLEX THAT CONTAINS SIN3A/B
CC AND HISTONE DEACETYLASE HDAC1 AND HDAC2. THIS COMPLEX ASSOCIATES
CC WITH THE THYROID (TR) AND THE RETINOID ACID RECEPTORS (RAR) IN THE
CC ABSENCE OF LIGAND.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM: ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: UBQUITOUS.
CC -1- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT
CC ARE DIVIDED INTO THREE INDEPENDANT REPRESSION DOMAINS (RD1, RD2
CC AND RD3). THE C-TERMINAL REGION CONTAINS THE NUCLEAR RECEPTOR-
CC INTERACTING DOMAINS THAT ARE DIVIDED IN TWO SEPARATE RECEPTION
CC DOMAINS (ID1 AND ID2).
CC -1- DOMAIN: THE TWO INTERACTION DOMAINS (ID) CONTAIN A CONSERVED
CC SEQUENCE REFERRED TO AS THE CORNR BOX. THIS MOTIF IS REQUIRED AND
CC SUFFICIENT TO PERMIT BINDING TO UNLIGANDED TR AND RARS. SEQUENCES
CC FLANKING THE CORNR BOX DETERMINE NUCLEAR HORMONE RECEPTOR
CC SPECIFICITY.
CC -1- SIMILARITY: CONTAINS 1 SANT DOMAIN (COMPOSED OF SANT-A+SANT-B).
CC -1- SIMILARITY: CONTAINS 2 CORNR BOX.
CC -1- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)

```

```

CC or send an email to license@isb-sib.ch.
CC -----
DR EMBL; U35312; AAB17125.1; -
DR EMBL; U22016; AAC52168.1; -
DR MGD; MGI:1349717; Ncor1.
DR InterPro: IPR001005; Myb_DNA_bind.
DR Pfam; PF00249; myb_DNA-binding; 2.
DR SMART; SM00395; SANT; 2.
DR PROSITE; PS50090; MYB_3; 1.
KW Nuclear protein; Transcription regulation; DNA-binding; Repressor;
KW Coiled coil; Alternative splicing.
FT DOMAIN 174 216
FT 254 312 COILED COIL (POTENTIAL).
FT 299 328 INTERACTION WITH SIN3A/B (BY SIMILARITY).
FT DNA_BIND 437 482 COILED COIL (POTENTIAL).
FT DNA_BIND 624 669 SANT-A (POTENTIAL).
FT DOMAIN 501 550 SANT-B (POTENTIAL).
FT 606 616 COILED COIL (POTENTIAL).
FT 606 616 PRO-RICH.
FT 2073 2077 CORNR BOX OF ID1.
FT 2277 2281 CORNR BOX OF ID2.
FT 58 64 POLY-GLN.
FT 593 602 POLY-ALA.
FT 1044 1047 POLY-PRO.
FT 1713 1718 POLY-ALA.
FT 1968 1979 POLY-SER.
FT 2333 2371 MISSING (IN SHORT ISOFORM).
FT 1952 1952 I -> T (IN REF. 2).
FT 2090 2090 A -> P (IN REF. 2).
SQ SEQUENCE 2453 AA; 270640 MW; 52208B40382F7E6A CRC64;

Query Match      71.1%; Score 54; DB 1; Length 2453;
Best Local Similarity 80.0%; Pred. No. 82;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      4 PPPPPAPQ 13
        ||||| 1:
Db      606 PPPPPPEP 615

RESULT 10
NCRL_MOUSE
ID NCRL_MOUSE STANDARD: PRT: 3119 AA.
AC P42859;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HUNTINGTIN (HUNTINGTON'S DISEASE PROTEIN HOMOLOG) (HD PROTEIN).
CN HD OR HDH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain, and Spleen;
RX MEDLINE=94214482; PubMed=8162057;
RA Lin B., Nasir J., Macdonald H., Hutchinson G., Graham R.K.,
RA Rommens J.M., Hayden M.R.;
RT "Sequence of the murine Huntington disease gene: evidence for
RT conservation, alternate splicing and polymorphism in a triplet
RT repeat.";
RL Hum. Mol. Genet. 3:85-92(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94278649; PubMed=8009370;
RA Barnes G.T., Duyao M.P., Ambrose C.M., McNeil S., Persichetti F.,
RA Strindahl J., Gusella J.F., Macdonald M.E.;
RT "Mouse Huntington's disease gene homolog (Hdh).";
RL Somat. Cell Mol. Genet. 20:87-97(1994).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=95375771; PubMed=7647777;

```

RA Trotter Y., Devys D., Imbert G., Saudou F., An I., Lutz Y.,  
 RA Weber C., Agid Y., Hirsch E.C., Mandel J.L.;  
 RT "Cellular localization of the Huntington's disease protein and  
 RT discrimination of the normal and mutated form."; Nat. Genet. 10:104-110(1995).  
 RL [4]  
 RN  
 RP SEQUENCE OF 1-181 FROM N.A.  
 RX MEDLINE=95278941; PubMed=7759106;  
 RA Lin B., Nasir J., Kaichman M.A., McDonald H., Zeisler J.,  
 RA Goldberg Y.P., Hayden M.R.;  
 RT "Structural analysis of the 5' region of mouse and human Huntington  
 RT disease genes reveals conservation of putative promoter region and  
 RT di- and trinucleotide polymorphisms."; Genomics 25:707-715(1995).  
 RL  
 CC -1- FUNCTION: MAY PLAY A ROLE IN MICROTUBULE-MEDIATED TRANSPORT OR  
 CC VESICLE FUNCTION.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -1- ALTERNATIVE PRODUCTS: ALTERNATIVE SPLICING SEEMS TO GIVE RISE TO A  
 CC 480 AMINO ACIDS SHORTER ISOFORM. THIS PRODUCT CANNOT BE EXPLAINED  
 CC BY A SIMPLE SPLICING EVENT.  
 CC -1- TISSUE SPECIFICITY: THE HIGHEST LEVEL IS SEEN THROUGHOUT THE  
 CC BRAIN, BUT IT IS ALSO FOUND IN THE STOMACH, HEART, TESTIS, ADIPOSE  
 CC TISSUE, MUSCLE, SPLEEN, LIVER, AND KIDNEY.  
 CC -1- DEVELOPMENTAL STAGE: PREDOMINANT EXPRESSION IN NEURONAL TISSUES AT  
 CC ALL DEVELOPMENTAL STAGES. IN 14.5 DAY OLD EMBRYOS, IT IS ALSO  
 CC DETECTED IN NON-NEURONAL TISSUES. THIS EXPRESSION IS DOWN-  
 CC REGULATED IN LATER STAGES OF DEVELOPMENT.  
 CC -1- POLYMORPHISM: THE FIRST POLY-PRO REPEAT STRETCH DIFFERS IN LENGTH  
 CC BY ONE UNIT (THREE) IN MUS SPRETUS STRAIN COMPARED TO OTHER  
 CC STRAINS (FOUR). THE POLY-GLN REGION DOES NOT APPEAR TO BE  
 CC POLYMORPHIC, EXPLAINING THE ABSENCE OF A MURINE HD-LINE DISORDER.  
 CC -1- SIMILARITY: CONTAINS 10 HEAT REPEATS.  
 CC -1- SIMILARITY: BELONGS TO THE HUNTINGTIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL; L23312; AAA37799.1; -  
 CC EMBL; L23313; AAA37800.1; -  
 CC EMBL; L28827; AAA89100.1; ALT. SEQ.  
 CC EMBL; U24233; AAC52218.1; -  
 CC EMBL; U34024; AAA91085.1; -  
 CC DR EMBL; U34008; AAA91085.1; JOINED.  
 CC DR EMBL; U34021; AAA91085.1; JOINED.  
 CC DR EMBL; U34022; AAA91085.1; JOINED.  
 CC DR EMBL; U34023; AAA91085.1; JOINED.  
 CC DR EMBL; MGI:96067; Hdh.  
 CC MGD; MGI:96067; Hdh.  
 CC InterPro: IPR000091; Huntingtin.  
 DR PRINTS: PR00375; HUNTINGTIN.  
 DR  
 KW Alternative splicing; Polymorphism; Repeat.  
 FT DOMAIN 182 306  
 FT DOMAIN 723 920  
 FT DOMAIN 1512 1553  
 FT DOMAIN 18 24  
 FT DOMAIN 25 45  
 FT DOMAIN 49 59  
 FT DOMAIN 1417 1420  
 FT DOMAIN 1696 1699  
 FT DOMAIN 2615 2620  
 FT DOMAIN 1522 2001  
 FT VARSPLIC 2 2  
 FT CONFLICT 29 29  
 FT CONFLICT 116 116  
 FT CONFLICT 138 138  
 FT CONFLICT 521 521  
 FT CONFLICT 524 524  
 FT CONFLICT 533 533  
 FT CONFLICT 607 607  
 A -> G (IN REF. 1 AND 4).  
 A -> P (IN REF. 2).  
 M -> L (IN REF. 2 AND 4).  
 N -> D (IN REF. 1).  
 S -> P (IN REF. 1).  
 A -> P (IN REF. 1).  
 A -> P (IN REF. 1).  
 A -> T (IN REF. 1).

FT CONFLICT 769 769 D -> E (IN REF. 2).  
 FT CONFLICT 972 972 S -> R (IN REF. 1).  
 FT CONFLICT 1106 1106 W -> C (IN REF. 1).  
 FT CONFLICT 1240 1240 T -> N (IN REF. 1).  
 FT CONFLICT 1384 1384 N -> T (IN REF. 1).  
 FT CONFLICT 1827 1827 H -> Y (IN REF. 1).  
 FT CONFLICT 1979 1980 H -> Y (IN REF. 1).  
 FT CONFLICT 2062 2062 PF -> SS (IN REF. 1).  
 FT CONFLICT 2570 2570 D -> G (IN REF. 1).  
 FT CONFLICT 2866 2866 S -> N (IN REF. 1).  
 FT CONFLICT 2877 2877 E -> V (IN REF. 1).  
 FT CONFLICT 2882 2882 V -> G (IN REF. 1).  
 FT CONFLICT 2887 2887 Q -> G (IN REF. 1).  
 FT CONFLICT 2915 2915 O -> H (IN REF. 1).  
 FT CONFLICT 3025 3025 A -> T (IN REF. 1).  
 FT CONFLICT 3025 3025 P -> S (IN REF. 3).  
 FT CONFLICT 3062 3063 OV -> LM (IN REF. 1).  
 FT CONFLICT 3095 3096 VV -> EE (IN REF. 1).  
 SQ SEQUENCE 3119 AA; 344685 MW; ECA42B5916F50F4F CRC64;  
 QY 3 APPPPPPAPQ 13  
 Db 29 APPPPPPPPQ 39  
 Query Match 71.1%; Score 54; DB 1; Length 3119;  
 Best Local Similarity 81.8%; Pred. No. 1e+02;  
 Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 RESULT 11  
 ID CCTL\_MOUSE STANDARD; PRT; 724 AA.  
 AC 09QWV9; 0920U7;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE CYCLIN T1 (CYCLIN T) (CYCT1).  
 GN CCNT1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/C;  
 RA MEDLINE=99059742; PubMed=9843510;  
 RA Bieniasz P.D., Girdina T.A., Bogerd H.P., Cullen B.R.;  
 RT "Recruitment of a protein complex containing Tat and cyclin T1 to TAR  
 RT governs the species specificity of HIV-1 Tat."; EMBO J. 17:7056-7065(1998).  
 RL EMBO J. 17:7056-7065(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Brain;  
 RA MEDLINE=99051315; PubMed=9832504;  
 RA Garber M.E., Wei P., KewalRamani V.N., Mayall T.P., Herrmann C.H.,  
 RA Rice A.P., Littman D.R., Jones K.A.;  
 RT "The interaction between HIV-1 Tat and human cyclin T1 requires zinc  
 RT and a critical cysteine residue that is not conserved in the murine  
 RT cyclin protein."; Genes Dev. 12:3512-3527(1998).  
 RL Genes Dev. 12:3512-3527(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A., AND MUTAGENESIS.  
 RC TISSUE=Spleen;  
 RX MEDLINE=99263519; PubMed=10329126;  
 RA Kwak Y.T., Ivanov D., Guo J., Nee E., Gaynor R.B.;  
 RT "Role of the human and murine cyclin T proteins in regulating HIV-1  
 RT Tat-activation."; J. Mol. Biol. 288:57-69(1999).  
 RL J. Mol. Biol. 288:57-69(1999).  
 RN [4]  
 RP SEQUENCE FROM N.A., AND MUTAGENESIS.  
 RC TISSUE=Fibroblast;  
 RX MEDLINE=99145542; PubMed=9990016;  
 RA Fujinaga K., Taube R., Wimmer J., Cujec T.P., Peterlin B.M.;



"Interactions between human cyclin T, Tat, and the transactivation response element (TAR) are disrupted by a cysteine to tyrosine substitution found in mouse cyclin T.";  
 Proc. Natl. Acad. Sci. U.S.A. 96:1285-1290(1999).  
 -1- FUNCTION: REGULATORY SUBUNIT OF THE CYCLIN-DEPENDENT KINASE PAIR (CDK9/CYCLIN T) COMPLEX, ALSO CALLED POSITIVE TRANSCRIPTION ELONGATION FACTOR B (P-TEFB), WHICH IS PROPOSED TO FACILITATE THE TRANSITION FROM ABORTIVE TO PRODUCTION ELONGATION BY PHOSPHORYLATING THE CTD (CARBOXY-TERMINAL DOMAIN) OF THE LARGE SUBUNIT OF RNA POLYMERASE II (RNAP II). DOES NOT BIND EFFICIENTLY TO THE TRANSCRIPTION DOMAIN OF THE HIV-1 NUCLEAR TRANSCRIPTIONAL ACTIVATOR, TAT.  
 -1- SUBUNIT: ASSOCIATES WITH CDK9 TO FORM P-TEFB. CYCLIN T1 IS THE PREDOMINANT CYCLIN ASSOCIATED WITH CDK9.  
 -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 -1- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. STRONG TO CYCLIN C SUBFAMILY.  
 -----  
 THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 -----  
 EMBL; AF095640; AAD13656.1; -  
 DR EMBL; AF109179; AAD19654.1; -  
 DR EMBL; AF087662; AAD1798.1; -  
 DR EMBL; AF113951; AAD17205.1; -  
 DR MGD; MGI:1328363; Ccml1.  
 DR InterPro; IPR000553; Cycl1n.  
 DR SMART; SM00385; CYCLIN; 1.  
 DR PROSITE; PS00292; CYCLINS; FALSE\_NEG.  
 KW Cyclin; Cell cycle; Cell division; Coiled coil;  
 TM Transcription regulation; Nuclear protein.  
 FT DOMAIN 253 270 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 254 272 TAT:TAR RECOGNITION MOTIF (TRM).  
 FT DOMAIN 319 322 POLY-SER.  
 FT DOMAIN 384 425 COILED COIL (POTENTIAL).  
 FT DOMAIN 516 525 POLY-HIS.  
 FT DOMAIN 565 569 POLY-SER.  
 FT DOMAIN 715 723 POLY-PRO.  
 FT CONFLICT 48 48 V -> M (IN REF. 3).  
 SQ SEQUENCE 724 AA; 80565 MW; 7396E428F5A8B91B CRC64;

Query Match 69.7%; Score 53; DB 1; Length 724;  
 Best Local Similarity 80.0%; Pred. No. 33;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 4 PPPPLPPAPQ 13  
 ||||| 1;  
 DB 715 PPPPLPPLPK 724

RESULT 12  
 CCT1\_HUMAN STANDARD; PRT; 726 AA.  
 AC 060563; 060581;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE CYCLIN T1 (CYCLIN T) (CCT1).  
 GN CCT1.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=T-cell lymphoma;  
 RX MEDLINE=96150851; PubMed=9491887;

RA Wei P., Garber M.E., Fang S.-M., Fischer W.H., Jones K.A.;  
 RT "A novel CDK9-associated C-type cyclin interacts directly with HIV-1  
 RT Tat and mediates its high-affinity, loop-specific binding to TAR  
 RT RNA.";  
 RL Cell 92:451-462(1998).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE=Breast cancer;  
 RX MEDLINE=98167917; PubMed=9499409;  
 RA Peng J.-M., Zhu Y., Milton J.T., Price D.H.;  
 RT "Identification of multiple cyclin subunits of human P-TEFB.";  
 RL Genes Dev. 12:755-762(1998).  
 RN [3]  
 RN FUNCTION.  
 RX MEDLINE=99263518; PubMed=10329125;  
 RA Ivanov D., Kwak Y.T., Nee E., Guo J., Garcia-Martinez L.F.,  
 RA Gaynor R.B.;  
 RT "Cyclin T1 domains involved in complex formation with Tat and TAR RNA  
 RT are critical for tat-activation.";  
 RL J. Mol. Biol. 288:41-56(1999).  
 RN [4]  
 RN MUTAGENESIS.  
 RX MEDLINE=99263519; PubMed=10329126;  
 RA Kwak Y.T., Ivanov D., Guo J., Nee E., Gaynor R.B.;  
 RT "Role of the human and murine cyclin T proteins in regulating HIV-1  
 RT Tat-activation.";  
 RL J. Mol. Biol. 288:57-69(1999).  
 RN [5]  
 RN FUNCTION.  
 RX MEDLINE=99051315; PubMed=9832504;  
 RA Garber M.E., Wei P., KewalRamani V.N., Mayall T.P., Herrmann C.H.,  
 RA Rice A.P., Littman D.R., Jones K.A.;  
 RT "The interaction between HIV-1 Tat and human cyclin T1 requires zinc  
 RT and a critical cysteine residue that is not conserved in the murine  
 RT Cycl1 protein.";  
 RL Genes Dev. 12:3512-3527(1998).  
 RN [6]  
 RN FUNCTION.  
 RX MEDLINE=99292873; PubMed=10364329;  
 RA Bieniasz P.D., Grdina T.A., Bogerd H.P., Cullen B.R.;  
 RT "Analysis of the effect of natural sequence variation in Tat and in  
 RT cyclin T on the formation and RNA binding properties of Tat-cyclin T  
 RT complexes.";  
 RL J. Virol. 73:5777-5786(1999).  
 CC -1- FUNCTION: REGULATORY SUBUNIT OF THE CYCLIN-DEPENDENT KINASE PAIR  
 CC (CDK9/CYCLIN T) COMPLEX, ALSO CALLED POSITIVE TRANSCRIPTION  
 CC ELONGATION FACTOR B (P-TEFB), WHICH IS PROPOSED TO FACILITATE THE  
 CC TRANSITION FROM ABORTIVE TO PRODUCTION ELONGATION BY  
 CC PHOSPHORYLATING THE CTD (CARBOXY-TERMINAL DOMAIN) OF THE LARGE  
 CC SUBUNIT OF RNA POLYMERASE II (RNAP II). BINDS ALSO TO THE  
 CC TRANSCRIPTIONAL DOMAIN OF THE HIV-1 NUCLEAR TRANSCRIPTIONAL  
 CC ACTIVATOR, TAT. IT IS PROBABLY THE CIS-ACTING COFACTOR FOR TAT. ALSO SERVES  
 CC AS AN ESSENTIAL COFACTOR FOR HIV-2-TAT AND SIMIAN IMMUNODEFICIENCY  
 CC VIRUS (AFRICAN MANDRILL) TAT FUNCTION.  
 CC -1- SUBUNIT: ASSOCIATES WITH CDK9 TO FORM P-TEFB. CYCLIN T1 IS THE  
 CC PREDOMINANT CYCLIN ASSOCIATED WITH CDK9.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- TISSUE SPECIFICITY: UBICITOUSLY EXPRESSED.  
 CC -1- MISCELLANEOUS: INTERACTION BETWEEN TAT AND CYCLIN T1 REQUIRES  
 CC ZINC.  
 CC -1- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. CYCLIN C SUBFAMILY.  
 -----  
 THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 -----  
 EMBL; AF045161; AAC39638.1; -  
 DR EMBL; AF048730; AAC39664.1; -;



```

DR MM; 602506;
DR InterPro; IPR000553; Cyclin.
DR SMART; SM00385; CYCLIN; 1.
DR PROSITE; PS00292; CYCLINS; FALSE_NEG.
KW Cyclin; Cell cycle; Cell division; Coiled coil;
KW Transcription regulation; Nuclear protein.
FT DOMAIN 253 270 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 254 272 TAT: TAR RECOGNITION MOTIF (TRM).
FT DOMAIN 384 425 COILED COIL (POTENTIAL).
FT DOMAIN 517 526 POLY-HIS.
FT DOMAIN 560 570 POLY-SER.
FT DOMAIN 717 725 POLY-PRO.
FT MUTAGEN 261 261 C->Y: LOSS OF HIV-1 TAT TRANSACTIVATION.
FT CONFLICT 77 77 Q -> R (IN REF. 2).
SQ SEQUENCE 726 AA; 80664 MW; 4637FEB2DDEFE13 CRC64;

Query Match
Best Local Similarity 69.7%; Score 53; DB 1; Length 726;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 4 PPPPLPPAPQ 13
DB 717 PPPPLPPLPK 726

RESULT 13
CCT1_HORSE
AC Q9XT26; STANDARD; PRT; 727 AA.
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE Cyclin T1 (CYCLIN T) (CCT1).
GN CCT1.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RS SEQUENCE FROM N.A.
RC TISSUE=Fibroblast;
RX MEDLINE=99303751; PubMed=10373508;
RA Bieniasz P.D., Grdina T.A., Bogerd H.P., Cullen B.R.;
RT "Highly divergent lentiviral Tat proteins activate viral gene
expression by a common mechanism."
RL Mol. Cell. Biol. 19:4592-4599(1999).
RT FUNCTION: REGULATORY SUBUNIT OF THE CYCLIN-DEPENDENT KINASE PAIR
(-1- CYCLIN T) COMPLEX. ALSO CALLED POSITIVE TRANSCRIPTON
ELONGATION FACTOR B (P-TEFB), WHICH IS PROPOSED TO FACILITATE THE
TRANSITION FROM ABORTIVE TO PRODUCTION ELONGATION BY
PHOSPHORYLATING THE CTD (CARBOXY-TERMINAL DOMAIN) OF THE LARGE
SUBUNIT OF RNA POLYMERASE II (RNAP II) (BY SIMILARITY). BINDS ALSO
TO THE TRANSACTIVATION DOMAIN OF THE EQUINE INFECTIOUS ANEMIA
VIRUS (EIAV) NUCLEAR TRANSCRIPTIONAL ACTIVATOR. TAT. IT IS
PROBABLY THE CIS-ACTING TRANSACTIVATION RESPONSE ELEMENT (TAR)
RNA-BINDING COFACTOR FOR TAT. DOES NOT BIND TO THE TRANSACTIVATION
DOMAIN OF THE HIV-1 NUCLEAR TRANSCRIPTIONAL ACTIVATOR.
CC -1- SUBUNIT: ASSOCIATES WITH CDK9 TO FORM P-TEFB. CYCLIN T1 IS THE
PREDOMINANT CYCLIN ASSOCIATED WITH CDK9.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. CYCLIN C SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC EMBL; AF137509; AAD38518.1; -
CC InterPro; IPR000553; Cyclin.

```

```

DR SMART; SM00385; CYCLIN; 1.
DR PROSITE; PS00292; CYCLINS; FALSE_NEG.
KW Cyclin; Cell cycle; Cell division; Coiled coil;
KW Transcription regulation; Nuclear protein.
FT DOMAIN 253 270 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 254 272 TAT: TAR RECOGNITION MOTIF (TRM) (BY
SIMILARITY).
FT DOMAIN 386 427 COILED COIL (POTENTIAL).
FT DOMAIN 519 528 POLY-HIS.
FT DOMAIN 562 573 POLY-SER.
FT DOMAIN 718 726 POLY-PRO.
SQ SEQUENCE 727 AA; 81013 MW; BFC2A398DB35BCE CRC64;

Query Match
Best Local Similarity 69.7%; Score 53; DB 1; Length 727;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 4 PPPPLPPAPQ 13
DB 718 PPPPLPPLPK 727

RESULT 14
FAST_HUMAN
ID FAST_HUMAN STANDARD; PRT; 281 AA.
AC P48023;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FAS ANTIGEN LIGAND (APOPTOSIS ANTIGEN LIGAND) (APTL).
OS Homo sapiens (Human).
GN TNFSF6 OR APTLIG1 OR FASL.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RS SEQUENCE FROM N.A.
RX MEDLINE=95105731; PubMed=7528780;
RA Alderson M.;
RT "Fas ligand mediates activation-induced cell death in human T
lymphocytes."
RL J. Exp. Med. 181:71-77(1995).
RN [2]
RS SEQUENCE FROM N.A.
RX MEDLINE=95127560; PubMed=7826947;
RA Takahashi T., Tanaka M., Inazawa J., Abe T., Suda T., Nagata S.;
RT "Human Fas ligand: gene structure, chromosomal location and species
specificity."
RL Int. Immunol. 6:1567-1574(1994).
RN [3]
RS SEQUENCE FROM N.A.
RX Schetzlein C.E., Poehlmann R., Philippson P., Eibel H.;
RA Submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.
RN [4]
RS SEQUENCE FROM N.A.
RX MEDLINE=95071350; PubMed=7980502;
RA Mita E., Hayashi N., Iio S., Takehara T., Hijioka T., Kasahara A.,
RA Fusamoto H., Kamada T.;
RT "Role of Fas ligand in apoptosis induced by hepatitis C virus
infection."
RL Biochem. Biophys. Res. Commun. 204:468-474(1994).
RN [5]
RS SEQUENCE FROM N.A.
RX Wilkison J.;
RA Submitted (MAY-1997) to the EMBL/Genbank/DBJ databases.
RN [6]
RS SEQUENCE OF 1-10 FROM N.A.
RC TISSUE=Blood;
RA Matsumura M., Nakanishi Y., Ohba Y.;
RL Submitted (APR-1998) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: CYTOKINE THAT BINDS TO FAS ANTIGEN, A RECEPTOR THAT
CC TRANSDUCES THE APOPTOTIC SIGNAL INTO CELLS. MAY BE INVOLVED IN

```







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2002, 16:30:02 ; Search time 40.9 Seconds  
(without alignments)  
46.492 Million cell updates/sec

Title: US-09-485-529-103  
Perfect score: 76  
Sequence: 1 LNAPPPPLPPAPQ 13

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues  
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP:archaea:\*  
2: SP:bacteria:\*  
3: SP:fungi:\*  
4: SP:human:\*  
5: SP:invertebrate:\*  
6: SP:mammal:\*  
7: SP:mhc:\*  
8: SP:organelle:\*  
9: SP:plant:\*  
10: SP:protist:\*  
11: SP:rodent:\*  
12: SP:virus:\*  
13: SP:vertebrate:\*  
14: SP:unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	76	100.0	623	10 Q9ST59	Q9ST59 tritlicum ae
2	59	77.6	180	5 Q9UB2	Q9UB2 leishmania
3	59	77.6	625	10 Q9MB96	Q9MB96 oryza sativ
4	57	75.0	282	4 Q9BSA1	Q9BSA1 homo sapien
5	57	75.0	407	2 Q9AAS9	Q9AAS9 caulobacter
6	57	75.0	862	5 Q44708	Q44708 drosophila
7	57	75.0	956	5 Q9GU21	Q9GU21 drosophila
8	57	75.0	1214	4 Q9BQ04	Q9BQ04 homo sapien
9	56	73.7	108	5 Q9VNZ6	Q9VNZ6 drosophila
10	56	73.7	391	2 Q9RFX5	Q9RFX5 bruceella su
11	56	73.7	599	10 P93797	P93797 volvox cart
12	56	73.7	630	10 Q9ST48	Q9ST48 zea mays (m
13	56	73.7	804	4 Q9UP56	Q9UP56 homo sapien
14	56	73.7	1943	11 Q9PFI1	Q9PFI1 mus musculu
15	55	72.4	267	5 Q18233	Q18233 caenorhabdi
16	55	72.4	369	11 Q9QYD1	Q9QYD1 mus musculu
17	55	72.4	572	11 Q9MX7	Q9MX7 mus musculu
18	55	72.4	579	10 Q9LGG8	Q9LGG8 oryza sativ
19	55	72.4	770	5 Q44014	Q44014 leishmania

20	55	72.4	1175	6 Q9TW66	Q9TW66 oryctolagus
21	54	71.1	67	11 Q9Z118	Q9Z118 rattus norv
22	54	71.1	138	10 Q9SSK0	Q9SSK0 arabidopsis
23	54	71.1	263	5 Q9VWK3	Q9VWK3 drosophila
24	54	71.1	302	5 Q9VND7	Q9VND7 drosophila
25	54	71.1	318	5 Q9VHT1	Q9VHT1 drosophila
26	54	71.1	375	11 Q9J35	Q9J35 mus musculu
27	54	71.1	451	11 Q62775	Q62775 rattus norv
28	54	71.1	485	11 Q92068	Q92068 rattus norv
29	54	71.1	780	10 Q9LKT8	Q9LKT8 arabidopsis
30	54	71.1	839	5 Q9Y103	Q9Y103 drosophila
31	54	71.1	839	5 Q9NMN7	Q9NMN7 drosophila
32	54	71.1	853	4 Q14333	Q14333 homo sapien
33	54	71.1	988	5 Q01864	Q01864 caenorhabdi
34	54	71.1	1043	5 Q9VUC6	Q9VUC6 drosophila
35	53	69.7	203	2 Q9RXC9	Q9RXC9 streptomyce
36	53	69.7	237	10 Q9FIP7	Q9FIP7 arabidopsis
37	53	69.7	325	10 Q9LRB7	Q9LRB7 oryza sativ
38	53	69.7	373	2 Q53253	Q53253 mycobacteri
39	53	69.7	385	5 Q9N5D3	Q9N5D3 caenorhabdi
40	53	69.7	409	10 Q9SBM1	Q9SBM1 volvox cart
41	53	69.7	420	13 Q9DEG1	Q9DEG1 gallus gall
42	53	69.7	507	10 Q9FU62	Q9FU62 oryza sativ
43	53	69.7	513	13 Q9DEG2	Q9DEG2 gallus gall
44	53	69.7	550	13 Q93263	Q93263 gallus gall
45	53	69.7	561	4 Q92952	Q92952 homo sapien

## ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	623 AA.
ID	Q9ST59			
AC	Q9ST59			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	GIBBERELLIN RESPONSE MODULATOR.			
GN	RHT-DIA.			
OS	Triticum aestivum (Wheat).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;			
OC	Triticeae; Triticum.			
OX	NCBI_TaxID=4565;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-99347734; PubMed-10421366;			
RA	Peng J., Richards D.E., Hartley N.M., Murphy G.P., Devos K.M.,			
RA	Flintham J.E., Beales J., Fish L.U., Worland A.D., Pelica F.,			
RA	Sudhakar D., Christou P., Snape J.W., Gale M.D., Harberd N.P.;			
RT	'Green revolution' genes encode mutant gibberellin response			
RT	modulators.";			
RL	Nature 400:256-261(1999).			
DR	EMBL: AJ242531; CAB51555.1; .			
SQ	SEQUENCE 623 AA; 65337 MW; F2BAC34996D0A84F CRC64;			

Query Match 100.0%; Score 76; DB 10; Length 623;  
Best local Similarity 100.0%; Pred. No. 0.013; Mismatches 0; Indels 0; Gaps 0;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1 LNAPPPPLPPAPQ 13	RESULT	2
DB	109 LNAPPPPLPPAPQ 121		
ID	Q9UB2	PRELIMINARY;	PRT; 180 AA.
AC	Q9UB2		
DT	01-MAY-2000 (TREMBLrel. 13, Created)		
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)		

DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE HYPOTHEITICAL 20.3 KDA PROTEIN.  
GN L2719.02.  
OS Leishmania major.  
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
OX NCBI\_TaxID=5664;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FRIEDLIN;  
RA Fuchs M., Gabel C., Mueller-Auer S., Schaefer M., Rieger M.,  
Ivens A.C., Lawson D., Quail M., Rajandream M.A., Barrell B.G.;  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FRIEDLIN;  
RX MEDLINE=98146435; PubMed=9477341;  
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,  
Smith D.F.;  
RA "A physical map of the Leishmania major Friedlin genome.";  
RL Genome Res. 8:135-145(1998).  
DR EMBL: AL117324; CAB55537.1; -.  
KW Hypothetical protein  
SQ SEQUENCE 180 AA; 20276 MW; 1F7E989F71E3A8CD CRC64;

Query Match 77.6%; Score 59; DB 5; Length 180;  
Best Local Similarity 90.9%; Pred. No. 0.69;  
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 LNAPPPPLPPA 11  
DB 83 LNAPPPPLPPA 93

RESULT 3  
OYMB96 PRELIMINARY; PRT; 625 AA.  
AC Q9MB96;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
DE OSGAI.  
GN OSGAI.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV, NIPPONBARE;  
RX MEDLINE=20179680; PubMed=10713441;  
RA Ogawa M., Kusano T., Katsumi M., Sano H.;  
"Rice gibberellin-insensitive gene homolog, OSGAI, encodes a nuclear-  
localized protein capable of gene activation at transcriptional  
level.";  
RL Gene 245:21-29(2000).  
DR EMBL: AB030956; BAA90749.1; -.  
SQ SEQUENCE 625 AA; 65406 MW; 034FF02719DA2E97 CRC64;

Query Match 77.6%; Score 59; DB 10; Length 625;  
Best Local Similarity 83.3%; Pred. No. 2.1;  
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 LNAPPPPLPPAP 12  
DB 110 LNAPPPPLPPAP 121

RESULT 4  
OYBSAI PRELIMINARY; PRT; 282 AA.

AC Q9BSAI;  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE ACTIVATING TRANSCRIPTION FACTOR 5.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=CHORIOCARCINOMA;  
RA Strausberg R.;  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC005174; AA05174.1; -.  
SQ SEQUENCE 282 AA; 30690 MW; B0D4CBF631655B1 CRC64;

Query Match 75.0%; Score 57; DB 4; Length 282;  
Best Local Similarity 100.0%; Pred. No. 1.9;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPAP 12  
DB 131 PPPPLPPAP 139

RESULT 5  
OYAA59 PRELIMINARY; PRT; 407 AA.  
AC Q9AA59;  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE OMPA FAMILY PROTEIN.  
GN CC0747.  
OS Caulobacter crescentus.  
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
OC Caulobacter.  
OX NCBI\_TaxID=69394;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21173698; PubMed=11259647;  
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
Pirochka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,  
Kolony J.F., Smit J., Craven M.B., Khouli H., Shetty J., Berry K.,  
Kuterback T., Tian K., Wolf A., Yamalhevyan J., Ermolaeva M., White O.,  
Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
"Complete genome sequence of Caulobacter crescentus.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
DR EMBL: AE005750; AAK22732.1; -.  
DR TIGR: CC0747; -.  
KW Complete proteome.  
SQ SEQUENCE 407 AA; 42781 MW; 54610D5FA64CC3BF CRC64;

Query Match 75.0%; Score 57; DB 2; Length 407;  
Best Local Similarity 100.0%; Pred. No. 2.6;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPAP 12  
DB 260 PPPPLPPAP 268

RESULT 6  
O44708 PRELIMINARY; PRT; 862 AA.  
AC O44708; Q9VE64; Q9VE65;  
DT 01-JUN-1998 (TREMBlrel. 06, Created)  
DT 01-JUN-2000 (TREMBlrel. 14, Last sequence update)

DT 01-JUN-2001 (TREMblrel. 17, last annotation update)  
 DE FRUITLESS PROTEIN, ISOFORMS MALE, FEMALE, 3 AND 4 (FRUITLESS CLASS I  
 GN FRU OR CG7688 OR CG7689 OR CG7690.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS MALE AND FEMALE).  
 RC STRAIN=OREGON-R;  
 RA MEDLINE=97133215; PubMed=8978612;  
 RA Ryner L.C., Goodwin S.F., Castellon D.H., Anand A., Vallelia A.,  
 RA Baker B.S., Hall J.C., Taylor B.J., Wasserman S.A.:  
 RT "Control of male sexual behavior and sexual orientation in Drosophila  
 by the fruitless gene".  
 RL Cell 87:1079-1089(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).  
 RC STRAIN=BERKELEY;  
 RA MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Chame M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,  
 RA Abrell J.F., Adyayani A., An H.-J., Andrews-Plankkuch C., Baldwin D.,  
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Bencos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos P., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasmo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Modary C., Morris J., Mostrel A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauley J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:  
 RT "The genome sequence of Drosophila melanogaster".  
 RL Science 287:2185-2195(2000).  
 CC -!- ALTERNATIVE PRODUCTS: 5 ISOFORMS: 3 (SHOWN HERE), 4, MALE, FEMALE  
 AND 5 (AC P91618), ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC EMBL: AF039231; AAB96677.1; -;  
 DR EMBL: AE003722; AAF5564.1; -;  
 DR EMBL: AE003722; AAF5565.1; ALT\_TERM.  
 DR EMBL: U72492; AAB92662.1; -;  
 DR FlyBase: FBgn0004652; fru.  
 DR InterPro: IPR000210; BTB\_POZ.  
 DR InterPro: IPR000361; EGF-like.  
 DR InterPro: IPR000822; Znf-C2H2.  
 DR Pfam: PF00651; BTB; 1.  
 DR Pfam: PF00096; zf-C2H2; 1.  
 DR SMART; SM00225; BTB; 1.

DR PROSITE; PS50097; BTB; 2.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 KW Alternative splicing, DNA-binding, EGF-like domain, Glycoprotein,  
 Zinc-finger.  
 FT VARSPLIC 1 101  
 FT MISSING (IN FEMALE ISOFORM AND ISOFORM  
 4).  
 FT MISSING (IN MALE ISOFORM AND FEMALE  
 ISOFORM).  
 FT RYKCNKIDRRPRDELNNRHHDDPCVIEVVDHYRE  
 MEAGNEHDEPKKEAHTATPPKGIKIGICLRDLR  
 VOQGAQHARHTMQSPADVDPLFAASAPPIPHRLPRGG  
 LITARGOQGLPGQOQSGAAGVFEHLPMRKPPTPAKIS  
 HTSOSLAEI -> SSARHILSTPLSTSSASPPPEFKHL  
 SALKREYHPLHYMANGGHNSALGYGNOGSGNAPNSAG  
 GAGSVAGVGAGGAGGATGAGGCHNHHTMSYHNMTPTSD  
 PGTMRRCRCGKEVYTRHHHSHRTORSCMCYCATYISRI  
 DTLSRLRVKHPDLRLKNSI (IN ISOFORM MALE  
 ISOFORM AND FEMALE ISOFORM).  
 FT MISSING (IN MALE ISOFORM AND FEMALE  
 ISOFORM).  
 FT VARSPLIC 617 769  
 FT MISSING (IN MALE ISOFORM AND FEMALE  
 ISOFORM).  
 FT VARSPLIC 790 862  
 FT MISSING (IN MALE ISOFORM AND FEMALE  
 ISOFORM).  
 SQ SEQUENCE 862 AA; 92593 MW; 94C229E91526701 CRC64;

Query Match Score 57; DB 5; Length 862;  
 Best Local Similarity 75.0%; Pred. No. 4.9;  
 Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 LNAPPPPLPPAP 12  
 Db 74 VRAPPPPLPPPP 85

RESULT 7  
 Q9GU21 PRELIMINARY; PRT; 956 AA.  
 AC Q9GU21;  
 DT 01-MAR-2001 (TREMblrel. 16, Created)  
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE FRUITLESS TYPE-A.  
 GN FRU OR CG7688 OR CG7689 OR CG7690.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CANTON S;  
 RA Usui-Aoki K., Ito H., Ui-Tel K., Takahashi K., Lukacovich T.,  
 RA Awano W., Nakata H., Piao Z.F., Nilsson E.E., Tomida J., Yamamoto D.:  
 RT "Formation of the male-specific muscle in female Drosophila by ectopic  
 fruitless expression".  
 RL Nat. Cell Biol. 0:0-0(2000).  
 DR EMBL: AF220177; AAC26588.1; -;  
 DR FlyBase: FBgn0004652; fru.  
 DR InterPro: IPR000210; BTB\_POZ.  
 DR InterPro: IPR000822; Znf-C2H2.  
 DR Pfam: PF00651; BTB; 1.  
 DR Pfam: PF00096; zf-C2H2; 1.  
 DR SMART; SM00225; BTB; 1.  
 DR SMART; SM00355; Znf-C2H2; 1.  
 DR PROSITE; PS50097; BTB; 1.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
 DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 1.  
 KW DNA-binding; Metal-binding; Zinc-finger.  
 SO SEQUENCE 956 AA; 103051 MW; 675260937C937FAF CRC64;

Query Match Score 57; DB 5; Length 956;  
 Best Local Similarity 75.0%; Pred. No. 5.4;  
 Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 LNAPPPPLPPAP 12  
: |||||  
DB 74 VRAPPPLPPPP 85

RESULT 8  
O9B004 PRELIMINARY: PRT: 1214 AA.

ID O9B004  
AC O9B004:  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DR KIAA0298 PROTEIN.  
GN KIAA0298.  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Amid C., Hanke T., Winterpacht A., Zabel B., Schmidt E.,  
RT "Comparative sequencing of human chromosome 11p15 and mouse chromosome  
RL Submitted (APR-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AJ400879; CAC35389.1;  
SQ SEQUENCE 1214 AA: 134403 MW: A3973CDEA2PD47C3A CRC64;

Query Match 75.0%; Score 57; DB 4; Length 1214;  
Best Local Similarity 90.0%; Pred. No. 6.7;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 PPPPLPPAPQ 13  
: |||||  
DB 465 PPPPLPPPPQ 474

RESULT 9  
O9VN26 PRELIMINARY: PRT: 108 AA.

ID O9VN26  
AC O9VN26:  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DE 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
GN CG14571 PROTEIN.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN-BERKELEY;  
RC MEDLINE=20196006; PubMed-10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Mead M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amaral L.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Brill J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borokova D., Botchan M.R., Bock J., Brooksstein P., Brotlier P.,  
RA Butlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Fodor J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin K.,  
RA Liu X., Maltel B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svrtkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of Drosophila melanogaster."  
RL Science 287:2185-2195(2000).  
DR EMBL; AF003595; AAF51766.1;  
DR FlyBase; FBgn0037118; CG14571.  
SQ SEQUENCE 108 AA: 11926 MW: B67F87AC2B2D3A45 CRC64;

Query Match 73.7%; Score 56; DB 5; Length 108;  
Best Local Similarity 90.0%; Pred. No. 1.1;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 APPPLPPAP 12  
: |||||  
DB 20 APPPLPPPPQ 29

RESULT 10  
O9RPX5 PRELIMINARY: PRT: 391 AA.

ID O9RPX5  
AC O9RPX5:  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
GN VIRB10.  
DE VIRB10.  
OS Brucella suis.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Brucellaceae; Brucella.  
OX NCBI\_TaxID=29461;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-1330;  
RX MEDLINE=99440171; PubMed-10510235;  
RA O'Callaghan D., Cazevielle C., Allardet-Servent A., Bour G.,  
RA Poulouange V., Frutos P., Kulakov Y., Ramuz M.,  
RT "A homologue of the Agrobacterium tumefaciens VirB and Bordetella  
RT pertussis Ptl Type IV secretion systems is essential for intracellular  
RT survival of Brucella suis."  
RL Mol. Microbiol. 33:1210-1220(1999).  
DR EMBL; AF141604; AAD56620.1;  
SQ SEQUENCE 391 AA: 41453 MW: A6DC4C9AFB8490D CRC64;

Query Match 73.7%; Score 56; DB 2; Length 391;  
Best Local Similarity 83.3%; Pred. No. 3.3;  
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 LNAPPPPLPPAP 12  
: |||||  
DB 81 LPAPPPPPPPAP 92

RESULT 11  
P93797 PRELIMINARY: PRT: 599 AA.



AC P93797;  
 DT 01-MAY-1997 (TREMBLrel. 03, Created)  
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE PHENOPHORIN-S PRECURSOR.  
 OS Volvox carterii.  
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 OC Volvocaceae; Volvox.  
 NCBI\_TaxID=3067;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-HK 10;  
 RX MEDLINE-97162277; PubMed-9009264;  
 RA Godl K., Hallmann A., Wenzl S., Sumper M.;  
 RT "Differential targeting of closely related ECM glycoproteins: the  
 RL pherophorin family from Volvox.";  
 EMBO J. 16:25-34(1997).  
 DR EMBL; Y07752; CAA69032.1; -;  
 DR Mendei; 14875; Volca;2392;14875.  
 KW Signal.  
 FT SIGNAL.  
 FT CHAIN 1 17 POTENTIAL.  
 FT SEQUENCE 599 AA; 63436 MW; 8BA1AD5A71F8082 CRC64;

Query Match  
 Best Local Similarity 73.7%; Score 56; DB 10; Length 599;  
 Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 NAPPPLPPAP 12  
 Db 216 NAPPSPLPSP 226

RESULT 12  
 ID 09ST48 PRELIMINARY; PRT; 630 AA.  
 AC 09ST48;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE GIBBERELLIN RESPONSE MODULATOR (FRAGMENT).  
 GN D8.  
 OS Zea mays (Maize).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
 OC Panicoideae; Andropogoneae; Zea.  
 NCBI\_TaxID=4577;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-99347734; PubMed-10421366;  
 RA Peng J., Richards D.E., Hartley N.M., Murphy G.P., Devos K.M.,  
 RA Filittham J.E., Beales J., Fish L.J., Worland A.J., Pelica F.,  
 RA Sudhakar D., Christou P., Snape J.W., Gale M.D., Harberd N.P.;  
 RT "Green revolution" genes encode mutant gibberellin response  
 RT modulators.";  
 RL Nature 400:256-261(1999).  
 DR EMBL; AJ242530; CAB51557.1; -;  
 FT NON\_TER 630 630  
 FT SEQUENCE 630 AA; 66028 MW; 3D56851726C51042 CRC64;

Query Match  
 Best Local Similarity 73.7%; Score 56; DB 10; Length 630;  
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 LNAPPPLPPA 11  
 Db 110 LNAPPAPLPAP 120

RESULT 13  
 Q9UP56

ID Q9UP56 PRELIMINARY; PRT; 804 AA.  
 AC Q9UP56;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE KIAA1076 PROTEIN (FRAGMENT).  
 GN KIAA1076.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RX MEDLINE-99397452; PubMed-10470851;  
 RA Kikuno R., Nagase T., Ishikawa K., Hirosewa M., Miyajima N.,  
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XIV.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro.";  
 RL DNA Res. 6:197-205(1999).  
 DR EMBL; AB028999; BA83028.1; -;  
 DR InterPro; IPR001214; SET.  
 DR InterPro; IPR003616; PostSET.  
 DR Pfam; PF00856; SET; 1.  
 DR PROSITE; PS50280; SET; 1.  
 DR SMART; SM00508; PostSET; 1.  
 DR SMART; SM00317; SET; 1.  
 FT NON\_TER 1  
 FT SEQUENCE 804 AA; 87997 MW; 13CAB0BA5420BE67 CRC64;

Query Match  
 Best Local Similarity 73.7%; Score 56; DB 4; Length 804;  
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 APPPLPPAP 12  
 Db 419 APPPLPPAP 428

RESULT 14  
 ID 099PJ1 PRELIMINARY; PRT; 1943 AA.  
 AC 099PJ1;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE PROTOCADHERIN.  
 GN AV.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-20578757; PubMed-11138007;  
 RA Alagramam K.N., Murcia C.L., Kwon H.Y., Pawlowski K.S., Wright C.G.,  
 RA Woychik R.P.;  
 RT "The mouse ames waltzer hearing-loss mutant is caused by mutation of  
 RT pcdh15, a novel protocadherin gene.";  
 RL Nat. Genet. 27:99-102(2001).  
 DR EMBL; AF281899; AAG53891.1; -;  
 FT SEQUENCE 1943 AA; 214816 MW; E3D7DB9F5F738652 CRC64;

Query Match  
 Best Local Similarity 73.7%; Score 56; DB 11; Length 1943;  
 Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 LNAPPPLPPAPQ 13  
 Db 1801 LSLPPPPPPAPR 1813

RESULT 15

018233

ID 018233 PRELIMINARY: PRT: 267 AA.

AC 018233;

DT 01-NOV-1996 (TREMblrel. 01, Created)

DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)

DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)

DE SIMILARITY WITH WILMS' TUMOR PROTEIN.

GN

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RN

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RX MEDLINE=94150718; PubMed=7906398;

RA Wilson R., Alnscough R., Anderson K., Baynes C., Berks M.,

RA Bonfield J., Burton J., Connell M., Copesey T., Cooper J., Coulson A.,

RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,

RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,

RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,

RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,

RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkhen R.,

RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,

RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,

RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RT elegans.";

RL Nature 368:32-38(1994).

RN

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA Geisel C., Steilyes L., Bradshaw H.;

RL Submitted (Apr-1996) to the EMBL/Genbank/DBJ databases.

RN

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RL Submitted (Mar-1996) to the EMBL/Genbank/DBJ databases.

RN

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RL Submitted (Nov-1996) to the EMBL/Genbank/DBJ databases.

DR EMBL; U53148; AAB37073.1; -;

SQ SEQUENCE 267 AA; 28592 MW; 039A28642CB6F175 CRC64;

Query Match 72.4%; Score 55; DB 5; Length 267;

Best Local Similarity 90.0%; Pred. No. 3.2;

Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 PPPPLPPAPQ 13

DB 90 PPPPPPPAPQ 99

Search completed: January 22, 2002, 16:30:03  
Job time: 338 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2002, 16:26:20 ; Search time 43.62 seconds

(without alignments)  
1057.947 Million cell updates/sec

Title: US-09-485-529-7

Perfect score: 623

Sequence: 1 MKREYDAGSGSGGGGSGMS.....TIGWHTPLATISAMRIAGP 623

Scoring table:

OLIGO  
Gapop 60.0 , Gapept 60.0

Searched: 522463 seqs, 74073290 residues

Word size : 0

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

A.Geneseq\_1101.\*  
1: /SID58/gcgdata/geneseq/geneseqp/AA1980.DAT.\*  
2: /SID58/gcgdata/geneseq/geneseqp/AA1981.DAT.\*  
3: /SID58/gcgdata/geneseq/geneseqp/AA1982.DAT.\*  
4: /SID58/gcgdata/geneseq/geneseqp/AA1983.DAT.\*  
5: /SID58/gcgdata/geneseq/geneseqp/AA1984.DAT.\*  
6: /SID58/gcgdata/geneseq/geneseqp/AA1985.DAT.\*  
7: /SID58/gcgdata/geneseq/geneseqp/AA1986.DAT.\*  
8: /SID58/gcgdata/geneseq/geneseqp/AA1987.DAT.\*  
9: /SID58/gcgdata/geneseq/geneseqp/AA1988.DAT.\*  
10: /SID58/gcgdata/geneseq/geneseqp/AA1989.DAT.\*  
11: /SID58/gcgdata/geneseq/geneseqp/AA1990.DAT.\*  
12: /SID58/gcgdata/geneseq/geneseqp/AA1991.DAT.\*  
13: /SID58/gcgdata/geneseq/geneseqp/AA1992.DAT.\*  
14: /SID58/gcgdata/geneseq/geneseqp/AA1993.DAT.\*  
15: /SID58/gcgdata/geneseq/geneseqp/AA1994.DAT.\*  
16: /SID58/gcgdata/geneseq/geneseqp/AA1995.DAT.\*  
17: /SID58/gcgdata/geneseq/geneseqp/AA1996.DAT.\*  
18: /SID58/gcgdata/geneseq/geneseqp/AA1997.DAT.\*  
19: /SID58/gcgdata/geneseq/geneseqp/AA1998.DAT.\*  
20: /SID58/gcgdata/geneseq/geneseqp/AA1999.DAT.\*  
21: /SID58/gcgdata/geneseq/geneseqp/AA2000.DAT.\*  
22: /SID58/gcgdata/geneseq/geneseqp/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	623	100.0	623	20	AAV02540
2	425	68.2	425	20	AAV02539
3	132	21.2	630	20	AAV02541
4	71	11.4	138	20	AAV02544
5	29	4.7	256	20	AAV02538
6	23	3.7	123	20	AAV02543
7	19	3.0	259	18	AAW30794
8	19	3.0	262	18	AAW38194
9	19	3.0	277	18	AAW38193
10	19	3.0	282	18	AAW30795
11	19	3.0	517	21	AAW38577

12	19	3.0	518	21	AAW38576	Arabidopsis thaliana
13	19	3.0	531	21	AAW28575	Arabidopsis SCLA2.
14	19	3.0	532	18	AAW30792	Arabidopsis thaliana
15	19	3.0	532	22	AAE02545	A. thaliana trans
16	19	3.0	532	22	AAE01907	Arabidopsis thaliana
17	19	3.0	533	21	AAW38575	Arabidopsis thaliana
18	19	3.0	572	22	AAW31883	Amino acid sequenc
19	19	3.0	572	22	AAW31884	AAW31884
20	19	3.0	587	21	AAW28574	Arabidopsis SCLA8.
21	19	3.0	587	22	AAE02560	A. thaliana trans
22	19	3.0	587	22	AAE01892	Arabidopsis thaliana
23	17	2.7	17	20	AAV02537	Peptide derived fr
24	17	2.2	100	20	AAV02542	Protein encoded by
25	14	2.2	221	19	AAW30793	Arabidopsis thaliana
26	13	2.1	1010	18	AAW34989	Arabidopsis thaliana
27	12	1.9	113	21	AAW32809	Teredinibacter end
28	12	1.9	833	21	AAW41835	EucaIyptus grandis
29	12	1.9	1136	20	AAV23330	Human OREF ORF1599
30	12	1.9	1136	20	AAV22655	Human tumour suppr
31	11	1.8	35	21	AAW11032	Human tumour thalia
32	11	1.8	216	21	AAW24090	Arabidopsis thaliana
33	11	1.8	216	21	AAW37177	Arabidopsis thaliana
34	11	1.8	231	22	AAW84954	Shrimp white spot
35	11	1.8	407	21	AAW06085	Arabidopsis thaliana
36	11	1.8	430	21	AAW17935	Arabidopsis thaliana
37	11	1.8	621	22	AAW84929	Shrimp white spot
38	11	1.8	797	20	AAV05850	Banana ripening fr
39	10	1.6	19	22	AAW17602	Peptide #4036 enco
40	10	1.6	19	22	AAW30122	Peptide #4159 enco
41	10	1.6	19	22	AAW05262	Peptide #3944 enco
42	10	1.6	25	22	AAW35702	Peptide #9739 enco
43	10	1.6	30	22	AAW21772	Peptide #8206 enco
44	10	1.6	30	22	AAW38096	Peptide #12153 enc
45	10	1.6	52	22	AAW32328	Peptide #6365 enco

#### ALIGNMENTS

RESULT 1	
AAV02540	AAV02540 standard; Protein; 623 AA.
XX	XX
AC	AAV02540;
XX	XX
DT	16-JUL-1999 (first entry)
XX	XX
DE	Protein encoded by wheat Rht clone 5a1 genomic sequence.
XX	XX
KW	Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;
KW	antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;
KW	paclobutrazol.
XX	XX
OS	Triticum aestivum.
XX	XX
PN	WO9909174-A1.
XX	XX
PD	25-FEB-1999.
XX	XX
PF	07-AUG-1998; 98WO-GB02383.
XX	XX
PR	13-AUG-1997; 97GB-0017192.
XX	XX
PA	(PLAN-) PLANT BIOSCIENCE LTD.
XX	XX
PI	Harberd NP, Peng J, Richards DE;
XX	XX
DR	WPI; 1999-181040/15.
XX	XX
PT	N-PSDB; AAW36279.
PT	New Triticum aestivum polynucleotides - encode a polypeptide which
PT	provides inhibition of the growth of plants, which inhibition is
PT	antagonised by gibberellin, used to confer a dwarf phenotype

XX Disclosure; Fig 8b; 88pp; English.  
XX  
CC The specification describes polypeptides encoded by the Rht gene (and  
CC its homologues) that, when expressed in *Triticum aestivum*, inhibit  
CC growth of the plant. This growth inhibition is antagonised by  
CC gibberellin. The products can be used to provide Rht expression in  
CC plants, conferring a dwarf phenotype on a plant which is correctable  
CC by treatment with gibberellin. In addition, the products can be  
CC used to produce Rht mutant plants which are dwarfed compared with  
CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants  
CC may be made by knocking out Rht or the relevant homologue gene in  
CC the plant of interest. Plants may be made which are resistant to  
CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,  
CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds  
CC dwarf but let crop plants grow tall. The present sequence is encoded by  
CC the wheat Rht clone 3a1 genomic sequence.  
XX  
XX

Sequence 623 AA;

Query Match 100.0%; Score 623; DB 20; Length 623;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 623; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKREYQDAGSGGSGGSGSEDKMMVSAAGEGEVEDELLAALGYKYRASDMADVAOKL 60  
DB 1 mkreyqdgsgggsgggsggsgsedkmmvsaagaegevdellaalgykyraadmadvaqkl 60  
QY 61 EOLENAMGMGVGACAPDDSFATHLATDYVHYNPTDLSWVESKMSLBNAPPPPLPAP 120  
DB 61 eplenamgmgyvgaapddsfathlatdyvhyndtdlsswesmlselbnappplppap 120  
QY 121 OLMASTSTVPSGGYFPLPSPVDSSTVYLRPIPSAGATAPADLSADSVRDKRMRT 180  
DB 121 olmaaststvggyfplpsspvdssstvyalrpi spagatapadlsadsvrdkrmrt 180  
QY 181 GGSSTSSSSSSSSSSSGGARSSVVEAPVAAAANATPALPVVVDTOEGIRLVHALLA 240  
DB 181 ggsstsssssssssggarssvveapvaaaanatapalpvvvdtdgeagrlrvhalla 240  
QY 241 CAENVQENLSAAELVYQIPLLAASOGGAMKVAAYFGELARRVFRPRPDSSLLDA 300  
DB 241 caenvqenlsaaelvyqipllaasgggammkvaayfgealarrvfrfrppdssllda 300  
QY 301 AFADLLAHFVSCPYLKFPAHTNAQALILEAPAGCRHVVDGIRKOGMOPALQALAL 360  
DB 301 afadllahfvescpylkfahftangallaefagcrrvhvdfgikgmgwpallqalal 360  
QY 361 RFGPPSFRLLGVGPQPDDETDAEQVGMKLAQFAHTIRVDFQYRGVLAATLADLEPML 420  
DB 361 rfgppsfrrllgvppqpdetdalaqvwmkilaqfahtrirvdfqyrglvaatladlepml 420  
QY 421 QPEGEDNEPEEVYLAIVSVFEMHRLAOPGALKEVLTGVRAVRRRIYTVVQENHNSG 480  
DB 421 qpegednepeevyiaivsvfemhrlaopgalekvltgvraivrriyvtvveqeanhns 480  
QY 481 TFLDFTSLHYSTMPFSLLEGSGSGGSEVSSGAAAPAAAGTDQVWSEYVLCROJCN 540  
DB 481 tfldfstlhystmfsllegsgsggsevs gaaapaaagtdqvmseylcrqjcn 540  
QY 541 VVACGABRTERHETLGQWRNRLGNAGETVHLGSNAYKQASTLLALFAGDGYVEEKE 600  
DB 541 vvacgaerterhetlgqwrnrlgnagetvhlgsnaykqaatlalafagdgylveeke 600  
QY 601 GCLITGWHTRPLIATSAARLACP 623  
DB 601 gclitgwhtrpliatcawrlagp 623

RESULT 2  
AA02539  
ID AA02539 standard; Protein: 425 AA.

XX  
AC AA02539;  
XX  
DT 16-JUL-1999 (first entry)  
XX  
DE Protein encoded by the wheat Rht clone C15-1 cDNA sequence.  
XX  
KW Rht gene; homologue; *Triticum aestivum*; wheat; growth inhibition;  
KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;  
KW paclobutrazol.  
XX  
OS *Triticum aestivum*.  
XX  
PN MO9909174-A1.  
XX  
PD 25-FEB-1999.  
XX  
PF 07-AUG-1998; 98WO-GB02383.  
XX  
PR 13-AUG-1997; 97GB-0017192.  
XX  
PA (PLAN-) PLANT BIOSCIENCE LTD.  
XX  
PI Harberd NP, Peng J, Richards DE;  
XX  
DR WPI: 1999-181040/15.  
XX  
DR N-PSDB: AAX36278.

PT New *Triticum aestivum* polynucleotides - encode a polypeptide which  
PT provides inhibition of the growth of plants, which inhibition is  
PT antagonised by gibberellin, used to confer a dwarf phenotype  
PS  
PS Disclosure; Fig 7b; 88pp; English.

CC The specification describes polypeptides encoded by the Rht gene (and  
CC its homologues) that, when expressed in *Triticum aestivum*, inhibit  
CC growth of the plant. This growth inhibition is antagonised by  
CC gibberellin. The products can be used to provide Rht expression in  
CC plants, conferring a dwarf phenotype on a plant which is correctable  
CC by treatment with gibberellin. In addition, the products can be  
CC used to produce Rht mutant plants which are dwarfed compared with  
CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants  
CC may be made by knocking out Rht or the relevant homologue gene in  
CC the plant of interest. Plants may be made which are resistant to  
CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,  
CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds  
CC dwarf but let crop plants grow tall. The present sequence is encoded by  
CC the wheat Rht clone C15-1 cDNA sequence.  
XX  
XX

Sequence 425 AA;

Query Match 68.2%; Score 425; DB 20; Length 425;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 425; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 199 ARSSVVEAPVAAAANATPALPVVVDTOEGIRLVHALLACAAVOQENLSAAEALVK 258  
DB 1 arssvveapvaaaanatapalpvvvdtdgeagrlrvhalla caenvqenlsaaelvyk 60  
QY 259 QIPLLAASOGGAMKVAAYFGELARRVFRPRPDSSLLDAAPDLAHNFVESCPLYK 318  
DB 259 qipllaasgggammkvaayfgealarrvfrfrppdsslldaafadllahfvescpylk 318  
QY 319 FAHFTANQALILEAPAGCRHVVDGIRKOGMOPALQALALRPGPSFRLLTGVGPOP 378  
DB 319 fahftanqalileafagcrrvhvdfgikgmgwpallqalalrpgpsfrrlltgvppgp 378  
QY 379 DETDALQOVGMKLAQFAHTIRVDFQYRGVLAATLADLEPMLQPEGEDNEPEEVIAVN 438  
DB 379 detdalaqvwmkilaqfahtrirvdfqyrglvaatladlepmlqpegednepeevia 438  
QY 439 SVFEMHRLAOPGALKEVLTGVRAVRRRIYTVVQENHNSGTFPLDRFTESLHYSTMF 498

```

Db      241 |=====|
          |sylvemrlllaqpalakvlgvtravprivtvegeanhnsgflldrfesllystmd 300
OY      499 |=====|
          |SLEGGSSGGSPSVSSGAAPAACTDOVMSRYVLGRQICNVVAEGAERPRRHETLQ 558
Db      301 |=====|
          |slegssggpsvssgaapaaagctdvmsevyigrqcnvacegaerterhel19 360
OY      559 |=====|
          |WRNRLGNAGFEVYHLGSMYKQASTLLALFAGDGYKVEKEGCLTLGWHTRPLIATSAW 618
Db      361 |=====|
          |wnrignagfevchysnaykqastllalfagdgdykvekegcltlgwhtrpliatasw 420
OY      619 |=====|
          |RLAGP 623
Db      421 |=====|
          |rlagp 425

```

## RESULT 3

AA02541  
ID AAY02541 standard; Protein; 630 AA.

AC AAY02541;

DT 16-JUL-1999 (first entry)

DE Protein encoded by maize 1a1 genomic clone sequence.

KW Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;  
KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;  
KW paclobutrazol; maize.

OS Zea mays.

PN WO909174-A1.

PD 25-FEB-1999.

PF 07-AUG-1998; 98WO-GB02383.

PR 13-AUG-1997; 97GB-0017192.

PA (PLAN-) PLANT BIOSCIENCE LTD.

PI Harberd NP, Peng J, Richards DE;

DR WPI; 1999-181040/15.

DR N-PSDB; AAX36280.

PT New Triticum aestivum polynucleotides - encode a polypeptide which  
PT provides inhibition of the growth of plants, which inhibition is  
PT antagonised by gibberellin, used to confer a dwarf phenotype

PS Disclosure; Fig 9b; 88pp; English.

CC The specification describes polypeptides encoded by the Rht gene (and  
CC its homologues) that, when expressed in Triticum aestivum, inhibit  
CC growth of the plant. This growth inhibition is antagonised by  
CC gibberellin. The products can be used to provide Rht expression in  
CC plants, conferring a dwarf phenotype on a plant which is correctable  
CC by treatment with gibberellin. In addition, the products can be  
CC used to produce Rht mutant plants which are dwarfed compared with  
CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants  
CC may be made by knocking out Rht or the relevant homologous gene in  
CC the plant of interest. Plants may be made which are resistant to  
CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,  
CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds  
CC dwarf but let crop plants grow tall. The present sequence is encoded by  
CC the maize 1a1 genomic clone sequence.

SO Sequence 630 AA;

Query Match 21.2%; Score 132; DB 20; Length 630;  
Best Local Similarity 100.0%; Pred. No. 3.9e-105;

```

Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      293 |=====|
          |PDSLLDAAPADLHAHFYESCPLYKFAHTTANQALIEAFAGCRVHVVDFGIKGOMQ 352
Db      302 |=====|
          |pdsllidaafadllhahfescplykfahtanqalleafagcrvhwvdfgikgmqp 361
OY      353 |=====|
          |ALQALALRPGPPSRLTGVGPPDETDALQOVGKTLAQFAHTTRVDQYRGLVAATL 412
Db      362 |=====|
          |allqalalrpgppsfiltcvppgpdetdalqvgvkwlaqfahllrvdfqyglvaatl 421
OY      413 |=====|
          |ADLEPMLQPEG 424
Db      422 |=====|
          |adlepmlqpeg 433

```

## RESULT 4

AA02544  
ID AAY02544 standard; Protein; 138 AA.

AC AAY02544;

DT 16-JUL-1999 (first entry)

DE Protein encoded by the partial sequence of the wheat rht-10 allele.

KW Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;  
KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;  
KW paclobutrazol.

OS Triticum aestivum.

PN WO909174-A1.

PD 25-FEB-1999.

PF 07-AUG-1998; 98WO-GB02383.

PR 13-AUG-1997; 97GB-0017192.

PA (PLAN-) PLANT BIOSCIENCE LTD.

PI Harberd NP, Peng J, Richards DE;

DR WPI; 1999-181040/15.

PT New Triticum aestivum polynucleotides - encode a polypeptide which  
PT provides inhibition of the growth of plants, which inhibition is  
PT antagonised by gibberellin, used to confer a dwarf phenotype

PS Disclosure; Fig 12b; 88pp; English.

CC The specification describes polypeptides encoded by the Rht gene (and  
CC its homologues) that, when expressed in Triticum aestivum, inhibit  
CC growth of the plant. This growth inhibition is antagonised by  
CC gibberellin. The products can be used to provide Rht expression in  
CC plants, conferring a dwarf phenotype on a plant which is correctable  
CC by treatment with gibberellin. In addition, the products can be  
CC used to produce Rht mutant plants which are dwarfed compared with  
CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants  
CC may be made by knocking out Rht or the relevant homologous gene in  
CC the plant of interest. Plants may be made which are resistant to  
CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,  
CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds  
CC dwarf but let crop plants grow tall. The present sequence is encoded by  
CC the partial sequence of the wheat rht-10 allele.

SO Sequence 138 AA;

Query Match 11.4%; Score 71; DB 20; Length 138;  
Best Local Similarity 100.0%; Pred. No. 3.3e-53;  
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 122 LNASTSTVTGSGGYEDLPSPVSSSSIVALRPIPSAGATAPADLSADSVDRPKRMRTG 181
    |||||||
DB 68 LNAststvtgsggyfdlppsvdsssslyalrpi spagatatpadlsadsvdrpkrmrtg 127
OY 182 GSSTSSSSSS 192
    |||||||
DB 128 gsstssssss 138

RESULT 5
AA02538
ID AA02538 standard; Protein: 256 AA.
XX
AC AA02538;
XX
DT 16-JUL-1999 (first entry)
XX
DE Protein encoded by rice EST D39460 sequence.
XX
KW Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;
KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;
KW paclobutrazol; rice; expressed sequence tag; EST.
XX
OS Oryza sativa.
XX
PN MO9909174-A1.
XX
PD 25-FEB-1999.
XX
PF 07-AUG-1998; 98WO-GB02383.
XX
PR 13-AUG-1997; 97GB-0017192.
XX
PA (PLAN-) PLANT BIOSCIENCE LTD.
XX
PI Harberd NP, Peng J, Richards DE;
XX
DR WPI: 1999-181040/15.
DR N-PSDB; AAX36277.
XX
PT New Triticum Aestivum polynucleotides - encode a polypeptide which
PT provides inhibition of the growth of plants, which inhibition is
PT antagonised by gibberellin, used to confer a dwarf phenotype
XX
PS Claim 12; Fig 6b; 88bp; English.
XX
CC The specification describes polypeptides encoded by the Rht gene (and
CC its homologues) that, when expressed in Triticum Aestivum, inhibit
CC growth of the plant. This growth inhibition is antagonised by
CC gibberellin. The products can be used to provide Rht expression in
CC plants, conferring a dwarf phenotype on a plant which is correctable
CC by treatment with gibberellin. In addition, the products can be
CC used to produce Rht mutant plants which are dwarfed compared with
CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants
CC may be made by knocking out Rht or the relevant homologous gene in
CC the plant of interest. Plants may be made which are resistant to
CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,
CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds
CC dwarf but let crop plants grow tall. The present sequence is encoded by
CC rice expressed sequence tag (EST) AAD39460, which is homologous to the
CC wheat Rht gene.
XX
SQ Sequence 256 AA;

```

```

Query Match 4.7%; Score 29; DB 20; Length 256;
Best Local Similarity 100.0%; Pred. No. 7.9e-17;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 221 PUVVVDQEGAGIRLVHALLACAEVQDEN 249
    |||||||
DB 227 pvvvdteagagirlvhalacaeavqden 255

```

```

RESULT 6
AA02543
ID AA02543 standard; Protein: 123 AA.
XX
AC AA02543;
XX
DT 16-JUL-1999 (first entry)
XX
DE Protein encoded by the partial sequence of the maize D8-2023 allele.
XX
KW Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;
KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;
KW paclobutrazol; maize.
XX
OS Zea mays.
XX
PN MO9909174-A1.
XX
PD 25-FEB-1999.
XX
PF 07-AUG-1998; 98WO-GB02383.
XX
PR 13-AUG-1997; 97GB-0017192.
XX
PA (PLAN-) PLANT BIOSCIENCE LTD.
XX
PI Harberd NP, Peng J, Richards DE;
XX
DR WPI: 1999-181040/15.
DR N-PSDB; AAX36282.
XX
PT New Triticum Aestivum polynucleotides - encode a polypeptide which
PT provides inhibition of the growth of plants, which inhibition is
PT antagonised by gibberellin, used to confer a dwarf phenotype
XX
PS Disclosure; Fig 11d; 88bp; English.
XX
CC The specification describes polypeptides encoded by the Rht gene (and
CC its homologues) that, when expressed in Triticum Aestivum, inhibit
CC growth of the plant. This growth inhibition is antagonised by
CC gibberellin. The products can be used to provide Rht expression in
CC plants, conferring a dwarf phenotype on a plant which is correctable
CC by treatment with gibberellin. In addition, the products can be
CC used to produce Rht mutant plants which are dwarfed compared with
CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants
CC may be made by knocking out Rht or the relevant homologous gene in
CC the plant of interest. Plants may be made which are resistant to
CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,
CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds
CC dwarf but let crop plants grow tall. The present sequence is encoded by
CC the partial sequence of the maize D8-2023 allele.
XX
SQ Sequence 123 AA;

```

```

Query Match 3.7%; Score 23; DB 20; Length 123;
Best Local Similarity 100.0%; Pred. No. 6.1e-12;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 51 SDMDVADQKLEQLEMAMGMGVG 73
    |||||||
DB 35 sdmdvraqkleqlenamgmgyg 57

```

```

RESULT 7
AA030794
ID AA030794 standard; Protein: 259 AA.
XX
AC AA030794;
XX
DT 19-MAR-1998 (first entry)

```

```

OY 19-MAR-1998 (first entry)

```

DE	Arabidopsis thaliana gibberellin insensitivty gai-d2 gene product.
XX	
KW	Gibberellin insensitivity; gai-d2; plant growth inhibition;
KW	dwarf phenotype; lodging resistance; increased yield;
KW	flowering regulation; bolting inhibition; spinach; lettuce;
KW	antibody; identification; probe; primer; antisense; sense;
KW	expression regulation; co-suppression; rice;
KW	Bakane disease resistance.
XX	
OS	Arabidopsis thaliana.
PN	W09729123-A2.
PD	14-AUG-1997.
XX	
PF	12-FEB-1997; 97MO-GB00390.
XX	
PR	12-FEB-1996; 96GB-0002796.
XX	
PA	(INNE-) INNES CENT INNOVATIONS LTD JOHN.
PI	Carol P, Harberd NP, Peng J, Richards DE;
DR	WPJ: 1997-415295/38.
N-PSDB:	AAT91939.
XX	
PT	Nucleic acid encoding gibberellin inhibitor GAI and related
PT	antisense sequences - used to create tall, or particularly, dwarf
XX	plants, especially crops such as maize, rice and wheat
PS	Claim 17; Fig 6d; 76pp; English.
XX	
CC	The present sequence is the Arabidopsis thaliana gibberellin
CC	insensitivity (gai-d2) gene product (GAI-d2), the expression of
CC	which inhibits plant growth. However the inhibition is antagonised
CC	by gibberellin (GA), while gai expression confers a dwarf phenotype
CC	that is insensitive to GA. Manipulating gai-d2 and GAI-d2
CC	expression can produce tall or dwarf plants, particularly the
CC	latter for increased resistance to lodging and increased yield. It
CC	may also allow regulation of flowering, i.e. plants remain in the
CC	vegetative state until treated with GA, useful to inhibit bolting
CC	in spinach and lettuce. GAI-d2 can be used to raise specific
CC	antibodies for identifying homologous proteins or genes in other
CC	species. gai-d2 fragments can also be used as probes or primers to
CC	identify and clone related sequences, or in the preparation of
CC	antisense or sense expression regulating (co-suppressing)
CC	sequences. Rice plants that express GAI-d2 may be resistant to Bakane
CC	disease. Manipulation of gai-d2 and GAI-d2 makes it possible to
CC	tallor the degree of dwarfism and GA sensitivity to particular crops
CC	or situations.
XX	
SQ	Sequence 259 AA;
OY	314 CPYLKFAHFTANOALTEAF 332
ID	
DB	229 cpylkfahftanqalleaf 247
Query Match	3.0%; Score 19; DB 18; Length 259;
Best Local Similarity	100.0%; Pred. No. 3.2e-08;
Matches 19; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
RESULT	8
ID	AAW38194
AC	AAW38194 standard; Protein: 262 AA.
XX	
DT	11-MAY-1998 (first entry)
XX	
DE	Arabidopsis SCARECROW SRPaz protein.
XX	

KM	SCARECROW; SCR; SRP2; transgenic plant; root; gravitropism;
KW	crop improvement.
XX	
OS	Arabidopsis thaliana.
PN	WO9741152-A1.
PD	
PP	06-NOV-1997.
XX	
PF	25-APR-1997; 97WO-US07022.
XX	
PR	24-APR-1997; 97US-0842445.
PR	26-APR-1996; 96US-0638617.
XX	
PA	(UYNX ) UNIV NEW YORK STATE.
XX	
PI	Benfey PM, Dilaurenzio L, Helariutta Y, Malamy JE;
PI	Pysh L, Wysocka-Diller J;
DR	WPI; 1997-549683/50.
DX	
PT	DNA encoding Arabidopsis SCARECROW protein - useful to modify plant
PT	cell division and therefore alter root development, or alter plant
PT	stem or hypocotyl gravitropism
XX	
XX	Claim 2; Page 114-115; 221pp; English.
XX	
CC	This is the amino acid sequence of Arabidopsis SRP2, a protein
CC	that shows homology to Arabidopsis SCARECROW (SCR) protein (see
CC	AAW38178). SCR is a member of a novel protein family and contains a
CC	number of functional domains similar to those found in
CC	transcription factors. SCR is expressed specifically in embryo
CC	root progenitor tissue and in certain root and stem tissues. It
CC	regulates a specific asymmetric division, and controls gravitropic
CC	response in aerial structures and root formation. SCR proteins
CC	(see AAW38178-201) from dicots and monocots, including SCR proteins
CC	lacking 1-4 of MOTIFS I-VI, and polypeptides corresponding to MOTIF
CC	I, II, III, IV, V or VI of SCR are claimed. Transgenic plants can
CC	be engineered to overexpress the SCR protein, so that cell division
CC	is increased in roots, resulting in thicker root development, while
CC	a plant with an altered stem or hypocotyl gravitropism is less
CC	susceptible to lodging. Plants that contain an antisense molecule
CC	that suppresses the expression of endogenous SCR gene product have
CC	thinner root development (all claimed).
XX	
SQ	Sequence 262 AA;
XX	
Query Match	3.0%; Score 19; DB 18; Length 262;
Best Local Similarity	100.0%; Pred. No. 3.3e-08;
Matches 19; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	314 CPYLKFAHFTANQATLEAF 332
DB	123 cpylkfahftanqalleaf 141
XX	
RESULT 9	
ID	AAW38193
AAW38193	standard; Protein: 277 AA.
XX	
AC	AAW38193;
XX	
DT	11-MAY-1998 (first entry)
XX	
DE	Arabidopsis SCARECROW SRP2 protein.
XX	
KW	SCARECROW; SCR; SRP2; transgenic plant; root; gravitropism;
KW	crop improvement.
XX	
OS	Arabidopsis thaliana.
RH	Key Location/Qualifiers

```

FT Misc-difference 269 /note= "unidentified residue"
ET XX
FN PN W09741152-A1.
PD PD 06-NOV-1997.
PE PE 25-APR-1997; 97WO-US07022.
XX XX
PR PR 24-APR-1997; 97US-0842445.
PP PP 26-APR-1996; 96US-0638617.
PA PA (UYNV ) UNIV NEW YORK STATE.
XX XX
PI PI Benfey PN, Dilaurenzio L, Helariutta Y, Malamy JE;
PY PY Pysh L, Myssocka-Diller J;
DR DR WPI; 1997-549683/50.
XX XX
PT PT DNA encoding Arabidopsis SCARCCROW protein - useful to modify plant
PR PR cell division and therefore alter root development, or alter plant
PS PS stem or hypocotyl gravitropism
XX XX
PM PM Claim 2; Page 113-114; 221pp; English.
XX XX
PN PN This is the amino acid sequence of Arabidopsis SRPα8, a protein
PC PC that shows homology to Arabidopsis SCARCCROW (SCR) protein (see
CA CA AAW38178). SCR is a member of a novel protein family and contains a
CC CC number of functional domains similar to those found in
CT CT transcription factors. SCR is expressed specifically in embryo
CG CG root progenitor tissue and in certain root and stem tissues. It
CH CH regulates a specific asymmetric division, and controls gravitropic
CI CI response in aerial structures and root formation. SCR proteins
CM CM (see AAW3178-201) from dicots and monocots, including SCR proteins
CN CN lacking 1-4 of MOTIFS I-VI, and polypeptides corresponding to MOTIF
CO CO I, II, III, IV, V or VI of SCR are claimed. Transgenic plants can
CP CP be engineered to overexpress the SCR protein, so that cell division
CQ CQ is increased in roots, resulting in thicker root development, while
CR CR a plant with an altered stem or hypocotyl gravitropism is less
CS CS susceptible to lodging. Plants that contain an antisense molecule
CT CT that suppresses the expression of endogenous SCR gene product have
CU CU thinner root development (all claimed).
CV CV
CX CX
CY CY Sequence 277 AA;
DZ DZ
EA EA
EB EB
EC EC
ED ED
EE EE
EF EF
EG EG
EH EH
EI EI
EJ EJ
EK EK
EL EL
EM EM
EN EN
EO EO
EP EP
EQ EQ
ER ER
ES ES
ET ET
EU EU
EV EV
EW EW
EX EX
FY FY Arabidopsis thaliana.
FZ FZ

```

xx	WM9729123-A2.
xx	14-AUG-1997.
xx	12-FEB-1997; 97WO-GB00390.
xx	12-FEB-1996; 96GB-0002796.
xx	(INNE-) INNES CENT INNOVATIONS LTD JOHN.
xx	Carol P, Harberd NP, Peng J, Richards DE;
xx	WPI: 1997-415295/38.
xx	N-PSDB; AAT91940.
xx	Nucleic acid encoding gibberellin inhibitor GAI and related
xx	antisense sequences - used to create tall, or particularly, dwarf
xx	plants, especially crops such as maize, rice and wheat
xx	Claim 17; Fig 6f; 76pp; English.
xx	The present sequence is the Arabidopsis thaliana gibberellin
xx	insensitivity (gai-d3) gene product (GAI-d3), the expression of
xx	which inhibits plant growth. However the inhibition is antagonised
xx	by gibberellin (GA), while gai expression confers a dwarf phenotype
xx	that is insensitive to GA. Manipulating gai-d3 and GAI-d3
xx	expression can produce tall or dwarf plants, particularly the
xx	latter for increased resistance to lodging and increased yield. It
xx	may also allow regulation of flowering, i.e. plants remain in the
xx	vegetative state until treated with GA, useful to inhibit bolting
xx	in spinach and lettuce. GAI-d3 can be used to raise specific
xx	antibodies for identifying homologous proteins or genes in other
xx	species. gai-d3 fragments can also be used as probes or primers to
xx	identify and clone related sequences, or in the preparation of
xx	antisense or sense expression regulating (co-suppressing)
xx	sequences. Rice plants that express GAI-d3 may be resistant to Bakane
xx	disease. Manipulation of gai-d3 and GAI-d3 makes it possible to
xx	tailor the degree of dwarfism and GA sensitivity to particular crops
xx	or situations.
xx	Sequence 282 AA;
xx	Query Match 3.0%; Score 19; DB 18; Length 282;
xx	Best Local Similarity 100.0%; Pred. No. 3.5e-08;
xx	Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
xx	QY 314 CPYKFAHFTANQAILLEAF 332
xx	
xx	Db 229 CPYKFAHFTANQAILLEAF 247
xx	
xx	RESULT 11
xx	AAG38577 standard; Protein; 517 AA.
xx	AAG38577;
xx	18-OCT-2000 (first entry)
xx	Arabidopsis thaliana protein fragment SHQ ID NO: 47612.
xx	Protein identification: signal transduction pathway; metabolic pathway;
xx	hybridisation assay; genetic mapping; gene expression control; promoter;
xx	termination sequence.
xx	Arabidopsis thaliana.
xx	EP1033405-A2.
xx	06-SEP-2000.



PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145813.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-015138.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.

PR 28-SEP-1999; 99US-0166458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 3.0%; Score 19; DB 21; Length 517;  
Best Local Similarity 100.0%; Pred. No. 5,9e-08;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 314 CPYLKFAHFTANQALILEAF 332  
|||||  
Db 231 cpylkfahftangalleaf 249

## RESULT 12

ID AAG38576 standard; Protein; 518 AA.

XX AAG38576;

DT 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 47611.

KW Protein identification; signal transduction pathway; metabolic pathway;  
hybridisation assay; genetic mapping; gene expression control; promoter;  
termination sequence.

XX Arabidopsis thaliana.

XX EPI033405-A2.

PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 23-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140921.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.

PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0158293.  
PR 13-OCT-1999; 99US-0158294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

## Query Match 3.0%; Score 19; DB 21; Length 518;

Best Local Similarity 100.0%; Pred. No. 5.9e-08; Mismatches 0; Indels 0; Gaps 0;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 CPYLKFAHFTANQALIEAF 332  
|||||  
DB 232 CPYLKFAHFTANQALIEAF 250

RESULT 13  
AAB28575  
ID AAB28575 standard; Protein; 531 AA.

AC AAB28575;

DT 09-FEB-2001 (first entry)

DE Arabidopsis SCL2.

KW Scarecrow gene; SCR gene; plant; regulatory gene; SCARECROW-like; SCL;  
KW transgenic plant; cell division; molecular marker; herbicide resistance;  
KW salt resistance; pathogen resistance; insect resistance.

OS Arabidopsis thaliana.

PN WO200053723-A2.

PD 14-SEP-2000.

PF 07-MAR-2000; 2000MO-US05875.

PR 10-MAR-1999; 99US-0265585.

PA (UNIV ) UNIV NEW YORK STATE.

XX Bentley PN, Di Lorenzo L, Wysocka-Diller J, Malamy JE, Pysn L;

PI Helianthus y, Bruce W, Lim J;

XX WPI; 2000-594315/56.

PT Scarecrow gene useful for producing transgenic plants expressing genes  
PT whose product increases starch, lignin or cellulose biosynthesis and  
PT confers herbicide, pathogen or insect resistance -

```

XX Claim 14; Fig 13; 200pp; English.
PS
CC The present sequence is given in a specification relating to the
CC structure and function of a regulatory gene, SCARECROW (SCR).
CC SCARECROW-like (SCL) genes encoding proteins containing an amino acid
CC sequence similar to the sequence of MORF III (VHID) of Arabidopsis SCR
CC protein are disclosed. SCR and SCL genes are useful for producing
CC transgenic plants whose cell division is modified and root and/or stem
CC development and gravitropism of stem or hypocotyl is altered. Cell
CC division is increased or decreased in roots resulting in thicker or
CC thinner root development. The transgenic plants are useful for expressing
CC a gene of interest encoding a gene product that confers herbicide, salt,
CC pathogen or insect resistance in root or embryos and genes encoding
CC starch, lignin or cellulose biosynthesis in shoots. The SCR gene also
CC confers less susceptibility to lodging in the transgenic plants than a
CC wild-type plant. SCR gene sequences are also useful as molecular markers
CC for a quantitative trait e.g. root or gravitropism trait in molecular
CC breeding of crop plants.
XX
SQ Sequence 531 AA;

Query Match 3.0%; Score 19; DB 21; Length 531;
Best Local Similarity 100.0%; Pred. No. 6.1e-08;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 CPYLKFAHTANQALTEAF 332
DB 245 CPYLKfahftanqalleaf 263

RESULT 14
AAM30792
ID AAM30792 standard; Protein; 532 AA.
AC
XX AAM30792;
XX
DT 19-MAR-1998 (first entry)
XX
DE Arabidopsis thaliana gibberellin insensitivity gai gene product.
XX
KW Gibberellin insensitivity; gai; plant growth inhibition;
KW dwarf phenotype; lodging resistance; increased yield;
KW flowering regulation; bolting inhibition; spinach; lettuce;
KW antibody; identification; probe; primer; antisense; sense;
KW expression regulation; co-suppression; rice;
KW Bakane disease resistance.
XX
XX Arabidopsis thaliana.
OS
XX MO9729123-A2.
XX
XX 14-AUG-1997.
XX
XX 12-FEB-1997; 97WO-GB00390.
XX
XX 12-FEB-1996; 96GB-0002796.
XX
XX (INNE-) INNES CENT INNOVATIONS LTD JOHN.
XX
XX Carol P, Harberd NP, Peng J, Richards DE;
XX
XX WPI: 1997-415295/38.
XX
XX N-PSDB: AAT91937.
XX
XX Nucleic acid encoding gibberellin inhibitor GAI and related
XX antisense sequences - used to create tall, or particularly, dwarf
XX plants, especially crops such as maize, rice and wheat
XX
XX Claim 1; Fig 4; 76pp; English.
XX
XX The present sequence is the Arabidopsis thaliana gibberellin
CC

```

```

CC insensitivity (gai) gene product (GAI), the expression of which
CC inhibits plant growth. However the inhibition is antagonised by
CC gibberellin (GA), while gai expression confers a dwarf phenotype
CC that is insensitive to GA. Manipulating gai and GAI expression can
CC produce tall or dwarf plants, particularly the latter for increased
CC resistance to lodging and increased yield. It may also allow
CC regulation of flowering, i.e. plants remain in the vegetative state
CC until treated with GA, useful to inhibit bolting in spinach and
CC lettuce. GAI can be used to raise specific antibodies for
CC identifying homologous proteins or genes in other species. gai
CC fragments can also be used as probes or primers to identify and
CC clone related sequences, or in the preparation of antisense or
CC sense expression regulating (co-suppression) sequences. Rice plants
CC that express GAI may be resistant to Bakane disease. Manipulation
CC of gai and GAI makes it possible to tailor the degree of dwarfism
CC and GA sensitivity to particular crops or situations.
XX
SQ Sequence 532 AA;

Query Match 3.0%; Score 19; DB 18; Length 532;
Best Local Similarity 100.0%; Pred. No. 6.1e-08;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 CPYLKFAHTANQALTEAF 332
DB 246 CPYLKfahftanqalleaf 264

RESULT 15
AAE02545
ID AAE02545 standard; Protein; 532 AA.
AC
XX AAE02545;
XX
XX 10-AUG-2001 (first entry)
XX
DE A. thaliana transcription factor G308.
XX
KW Plant transcription factor; phenotype; sugar sensing characteristic;
KW transgenic plant; plant yield; growth; germination; photosynthesis;
KW glyoxylate metabolism; respiration; pathogen response; wounding response;
KW cell cycle regulation; pigmentation; flowering; senescence; physiology;
KW storage organ; metabolism.
XX
XX Arabidopsis thaliana.
OS
XX
XX Key Location/Qualifiers
XX Domain 270..274
XX /note="Conserved domain"

WO200135725-A1.
XX
XX 25-MAY-2001.
XX
XX 14-NOV-2000; 2000WO-US31414.
XX
XX 17-NOV-1999; 99US-0166228.
XX
XX 17-APR-2000; 2000US-0197899.
XX
XX 22-AUG-2000; 2000US-0227439.
XX
XX (MENDEL) MENDEL BIOTECHNOLOGY INC.
XX
XX (JIAN/) JIANG C.
XX
XX (HEAR/) HEARD J.
XX
XX (PINE/) PINEDA O.
XX
XX (PILG/) PILGRIM M.
XX
XX (ADAM/) ADAM L.
XX
XX (RIEC/) RIECHMANN J L.
XX
XX (YUGG/) YU G.
XX
XX (SAMA/) SAMAHA R.
XX
XX Jiang C, Heard J, Pineda O, Pilgrim M, Adam L, Riechmann JL;
XX Yu G, Samaha R;
PI

```

XX WPI: 2001-335977/35.  
DR N-PSDB; AAD06646.

XX Nucleic acids encoding plant transcription factor polypeptides, useful  
PT for altering the sugar sensing characteristics of plants and increasing  
PR yield, e.g. corn, potato and cotton plants -  
XX

PS Claim 4; Page 74-76; 151pp; English.

XX  
CC The patent relates to polynucleotides encoding 35 plant transcription  
CC factors which may be used to modify phenotype associated with a plant's  
CC sugar sensing characteristics and increasing yield when their expression  
CC level is altered. Sugars are central regulatory molecules that control  
CC aspects of physiology, metabolism and development. Therefore the cDNAs  
CC and proteins of the invention are useful for modifying the growth and  
CC germination rates of plants, photosynthesis, glyoxylate metabolism,  
CC respiration, starch and sucrose synthesis and degradation, pathogen  
CC response, wounding response, cell cycle regulation, pigmentation,  
CC flowering and senescence of plants and for modifying sink-source  
CC relationships in seeds, tubers, roots, and other storage organs leading  
CC to an increase in yield. The transcription factor polynucleotides and  
CC polypeptides may be used to alter the structure and developmental  
CC characteristics of plants such as soybean, wheat, corn, potato, cotton,  
CC rice, oilseed, sunflower, alfalfa, sugarcane, turf, banana, blackberry,  
CC blueberry, strawberry, raspberry, cantaloupe, carrot, cauliflower,  
CC coffee, cucumber, eggplant, grapes, honey dew, lettuce, mango, melon,  
CC onion, papaya, peas, peppers, pineapple, spinach, squash, sweet corn,  
CC tobacco, tomato, watermelon, rosaceous fruits and/or vegetable brassicas.  
CC The present sequence is an Arabidopsis thaliana transcription factor.  
XX

SO Sequence 532 AA;

Query Match 3.0%; Score 19; DB 22; Length 532;  
Best Local Similarity 100.0%; Pred. No. 6.1e-08;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 CPYIKFAHFTANQAIIEAF 332  
|||||  
DB 246 cpylkhfahftangalleaf 264

Search completed: January 22, 2002, 16:31:44  
Job time: 324 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2002, 16:29:20 ; Search time 26.82 seconds  
(without alignments)  
1769,454 Million cell updates/sec

Title: US-09-485-529-7

Perfect score: 623

Sequence: 1 MKREYDAGSGGGGGGSGMS.....TLGHTPRPLATSAWRLAGP 623

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	3.5	511	2	G96688
2	19	3.0	523	2	T51475
3	19	3.0	533	2	H86282
4	19	3.0	587	2	D84426
5	16	2.6	275	2	T04480
6	12	1.9	153	2	S67294
7	12	1.9	229	2	UC7219
8	12	1.9	390	2	E81408
9	12	1.9	446	2	T45525
10	12	1.9	475	2	S49886
11	11	1.8	164	2	T16168
12	11	1.8	216	2	B66794
13	11	1.8	238	2	T32889
14	11	1.8	429	2	E96685
15	11	1.8	575	2	T01552
16	11	1.8	632	2	T00679
17	11	1.8	693	2	S64904
18	11	1.8	816	2	S64439
19	11	1.8	1249	2	T62294
20	11	1.8	1772	2	A45532
21	10	1.6	205	2	T15450
22	10	1.6	210	2	T49785
23	10	1.6	230	2	T39598
24	10	1.6	238	2	T52505
25	10	1.6	240	2	D71272
26	10	1.6	242	2	D86343
27	10	1.6	246	2	S58497
28	10	1.6	262	2	T40941
29	10	1.6	265	2	D96667

30	10	1.6	284	2	S58650	hypothetical prote
31	10	1.6	306	2	T47832	hypothetical prote
32	10	1.6	349	2	T41394	hypothetical serin
33	10	1.6	406	2	T48103	mRNA binding prote
34	10	1.6	411	2	E96665	protein F22C12.16
35	10	1.6	412	2	T48405	myd-like protein -
36	10	1.6	414	2	D96838	unknown protein T2
37	10	1.6	421	2	S71331	L-ascorbate peroxi
38	10	1.6	430	2	T12282	L-ascorbate peroxi
39	10	1.6	443	2	T05540	hypothetical prote
40	10	1.6	461	2	T11819	glycerol-3-phospha
41	10	1.6	468	2	A55476	protein kinase (EC
42	10	1.6	484	2	S65713	hypothetical prote
43	10	1.6	503	2	S63257	probable membrane
44	10	1.6	506	2	F86733	ABC transporter pe
45	10	1.6	568	2	I39411	AF-9 protein - hum

#### ALIGNMENTS

```

RESULT 1
G96688.
hypothetical protein T27F4.10 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence-revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: G96688
R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.E.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Malt, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: G96688
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-511 <STO>
A:Cross-references: GB:AE005173; NID:q10092507; PIDN:AA612907.1; GSPDB:GM00141
A:Map position: 1
C:Genetics:
A:Gene: T27F4.10
Query Match 3.5%; Score 22; DB 2; Length 511;
Best Local Similarity 100.0%; Pred. No. 4.7e-11;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 309 HFEYSCPYLKFAHFTANQAIL 330
DB 225 HFEYSCPYLKFAHFTANQAIL 246
RESULT 2
T51475
RGA-like protein - Arabidopsis thaliana
N:Alternate names: protein K3M16.60
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 18-Aug-2000 #sequence-revision 18-Aug-2000 #text_change 18-Aug-2000
C:Accession: T51475
R:Satoh, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.;
submitted to the Protein Sequence Database, August 2000
A:Reference number: Z25394
A:Accession: T51475
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-523 <SAT>
A:Cross-references: EMBL:AL391150
A:Experimental source: cultivar Columbia; BAC clone K3M16

```

C:Genetics:  
A:Map position: 5  
A:Note: K3M16\_60

Query Match 3.0%; Score 19; DB 2; Length 523;  
Best Local Similarity 100.0%; Pred. No. 2e-08;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 313 CPYLKFAHFTANQALIEAF 331  
|||||  
Db 234 CPYLKFAHFTANQALIEAF 252

RESULT 3  
H86282  
Protein F10B6.34 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: H86282

R:Theologian, A.; Ecker, J.R.; Palm, C.J.; Federpspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Hultzar, L.  
Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719

A:Accession: H86282  
A:Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-533 <STO>  
A:Cross-references: GB:AE005172; NID:98778219; PIDN:AAF9228.1; GSPDB:GN00141

C:Genetics:  
A:Gene: F10B6.34  
A:Map position: 1

Query Match 3.0%; Score 19; DB 2; Length 533;  
Best Local Similarity 100.0%; Pred. No. 2e-08;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 CPYLKFAHFTANQALIEAF 332  
|||||  
Db 247 CPYLKFAHFTANQALIEAF 265

RESULT 4  
D84426  
hypothetical protein At2g01570 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: D84426

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffit, K.S.; Cronin, L.A.; Shen, M.; Vankken, S.E.; Umayam, L.; Tallon, L.  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487

A:Accession: D84426  
A:Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-587 <STO>  
A:Cross-references: GB:AE002093; NID:93785986; PIDN:AAC67333.1; GSPDB:GN00139

C:Genetics:  
A:Gene: At2g01570  
A:Map position: 2

Query Match 3.0%; Score 19; DB 2; Length 587;

Best Local Similarity 100.0%; Pred. No. 2.2e-08;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 CPYLKFAHFTANQALIEAF 332  
|||||  
Db 299 CPYLKFAHFTANQALIEAF 317

RESULT 5  
T04480  
acyl-CoA oxidase homolog - barley (fragment)  
C:Species: Hordeum vulgare (barley)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 08-Oct-1999  
C:Accession: T04480

R:Grossi, M.; Gull, M.; Stanca, A.M.; Cattiveill, L.  
Plant Sci. 105, 71-80, 1995

A:Title: Characterization of two barley genes that respond rapidly to dehydration str  
A:Reference number: 215371  
A:Accession: T04480  
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA  
A:Residues: 1-275 <GRC>  
A:Cross-references: EMBL:X84055; NID:9642243; PIDN:CA58874.1; PID:9642244

A:Experimental source: cv. Arda, leaf  
C:Genetics:  
A:Gene: cdt29

Query Match 2.6%; Score 16; DB 2; Length 275;  
Best Local Similarity 100.0%; Pred. No. 4.8e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 482 FLDRTESLHYSTMF 497  
|||||  
Db 8 FLDRTESLHYSTMF 23

RESULT 6  
S67294  
hypothetical protein YOR382w - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein O6760  
C:Species: Saccharomyces cerevisiae  
C:Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 05-Nov-1999  
C:Accession: S67294

R:Delius, H.; Hebling, U.; Hofmann, B.  
submitted to the Protein Sequence Database, July 1996

A:Reference number: S67261  
A:Accession: S67294  
A:Molecule type: DNA

A:Residues: 1-153 <DEU>  
A:Cross-references: EMBL:Z75290; NID:91420822; PID:e252449; PID:91420823; GSPDB:GN000

A:Experimental source: strain S288C  
C:Genetics:  
A:Gene: MIPS:YOR382w  
A:Map position: 15R

Query Match 1.9%; Score 12; DB 2; Length 153;  
Best Local Similarity 100.0%; Pred. No. 0.009;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 183 SSTSSSSSSSS 194  
|||||  
Db 111 SSTSSSSSSSS 122

RESULT 7  
JC7219  
nuclear protein SR-25 - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 21-Jul-2000  
C:Accession: JC7219

R:Sasahara, K.; Yamaoka, T.; Moritani, M.; Tanaka, M.; Iwahana, H.; Yoshimoto, K.; MI



Biochem. Biophys. Res. Commun. 269, 444-450, 2000  
A:Title: Molecular cloning and expression analysis of a putative nuclear protein, SR-25.  
A:Reference number: JG7219; MUID:20175222  
A:Accession: JG7219  
A:Molecule type: mRNA  
A:Residues: 1-229 <SAS>  
A:Cross-references: DDBJ:AB05383; NID:g7619895; PIDN:BAA94743.1; PID:g7619896  
A:Experimental source: MIN6 cell line  
C:Comment: This protein is a highly hydrophilic nuclear protein with a serine-arginine  
A splicing factors.  
C:Keywords: nucleus; RNA processing

Query Match 1.9%; Score 12; DB 2; Length 229;  
Best Local Similarity 100.0%; Pred. No. 0.013;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 183 SSTRSSSSSSSS 194  
|||||  
DB 69 SSTRSSSSSSSS 80

RESULT 8  
E81408  
probable periplasmic protein Cj0606 [imported] - Campylobacter jejuni (strain NCTC 11168  
C:Species: Campylobacter jejuni  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 31-Mar-2000  
C:Accession: E81408  
R:Parhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barre  
Nature 403, 665-668, 2000  
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp  
A:Reference number: AB1250; MUID:20150912  
A:Accession: E81408  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-390 <PAR>  
A:Cross-references: GB:AI139075; GB:AI111168; NID:g6967817; PIDN:CAB75242.1; PID:g696807  
A:Experimental source: serotype O2, strain NCTC 11168  
C:Genetics:  
A:Gene: Cj0606

Query Match 1.9%; Score 12; DB 2; Length 390;  
Best Local Similarity 100.0%; Pred. No. 0.02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 183 SSTRSSSSSSSS 194  
|||||  
DB 268 SSTRSSSSSSSS 279

RESULT 9  
T45525  
WSC4 homolog [imported] - Yeast (*Kluyveromyces marxianus* var. *lactis*)  
C:Species: *Kluyveromyces marxianus* var. *lactis*; *Candida spherica*  
C:Date: 31-Jan-2000 #sequence\_revision 31-Jan-2000 #text\_change 31-Jan-2000  
C:Accession: T45525  
R: Bao, W.G.; Fukuhara, H.  
submitted to the EMBL Data Library, July 1999  
A:Description: The ubiquitin-encoding genes of *Kluyveromyces lactis*.  
A:Reference number: Z23000  
A:Accession: T45525  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-446 <BAO>  
A:Cross-references: EMBL:AJ243800; PIDN:CAB50897.1  
A:Experimental source: strain 2359/152  
C:Genetics:  
A:Gene: wsc4

Query Match 1.9%; Score 12; DB 2; Length 446;

Best Local Similarity 100.0%; Pred. No. 0.022;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 183 SSTRSSSSSSSS 194  
|||||  
DB 184 SSTRSSSSSSSS 195

RESULT 10

S49886  
probable membrane protein Y1L123w - yeast (*Saccharomyces cerevisiae*)  
N:Alternate names: hypothetical protein Y18277.06  
C:Species: *Saccharomyces cerevisiae*  
C:Date: 13-Jan-1995 #sequence\_revision 20-Feb-1995 #text\_change 11-Jan-2000  
C:Accession: S49886  
R: Hamlyn, N.; Churcher, C.  
submitted to the EMBL Data Library, November 1994  
A:Reference number: S49881  
A:Accession: S49886  
A:Molecule type: DNA  
A:Residues: 1-475 <HAM>  
A:Cross-references: GB:Z47047; EMBL:Z46833; NID:g603997; PID:g763223; MIPS:Y1L123w  
C:Genetics:

A:Gene: SGD:SIM1  
A:Cross-references: SGD:S0001385; MIPS:Y1L123w  
A:Map position: 9L  
C:Superfamily: *Saccharomyces* NCA3 protein  
C:Keywords: transmembrane protein  
F:82-98/Domain: transmembrane #status predicted <TMM>

Query Match 1.9%; Score 12; DB 2; Length 475;  
Best Local Similarity 100.0%; Pred. No. 0.023;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 183 SSTRSSSSSSSS 194  
|||||  
DB 182 SSTRSSSSSSSS 193

RESULT 11  
T16168  
hypothetical protein F26F4.6 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: T16168  
R: Fulton, L.  
submitted to the EMBL Data Library, March 1996  
A:Description: The sequence of *C. elegans* cosmid F26F4.  
A:Reference number: Z18471  
A:Accession: T16168  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-164 <FUL>  
A:Cross-references: EMBL:U12964; NID:g1213452; PID:g529204; PIDN:AAA91221.1; CESP:F26  
A:Experimental source: strain Bristol N2  
C:Genetics:  
A:Gene: CESP:F26F4.6  
A:introns: 34/1

Query Match 1.8%; Score 11; DB 2; Length 164;  
Best Local Similarity 100.0%; Pred. No. 0.071;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 184 SSTRSSSSSSSS 194  
|||||  
DB 106 SSTRSSSSSSSS 116

RESULT 12  
B96794  
unknown protein F14G6.20 [imported] - *Arabidopsis thaliana*

C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: B96794  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpsiel, N.A.; Kaul, S.; White, O.; Alonso, C.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.; Dewar, K.; Jensen, N.F.; Hughes, B.; Hultzar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzalla, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719  
 A:Accession: B96794  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-216 <STO>  
 A:Cross-references: GB:AE005173; NID:96642677; PIDN:AAF20257.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: F14G6.20  
 A:Map position: 1

Query Match 1.8%; Score 11; DB 2; Length 216;  
 Best Local Similarity 100.0%; Pred. No. 0.089;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 184 STSSSSSSSS 194  
 Db 45 STSSSSSSSS 55

RESULT 13  
 T32889  
 hypothetical protein C34B2.9 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
 C:Accession: T32889  
 R:Graves, T.; Suterer, C.; Hawkins, M.; Wilson, R.  
 A:Description: The EMBL Data Library, January 1998  
 A:Reference number: 221241  
 A:Accession: T32889  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-238 <GBA>  
 A:Cross-references: EMBL:AF043693; PIDN:AAB97541.1; GSPDB:GN00019; CESP:C34B2.9  
 A:Experimental source: Strain Bristol N2; clone C34B2  
 C:Genetics:  
 A:Gene: CESP:C34B2.9  
 A:Map position: 1  
 A:Introns: 35/3; 91/2; 200/2

Query Match 1.8%; Score 11; DB 2; Length 238;  
 Best Local Similarity 100.0%; Pred. No. 0.096;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 184 STSSSSSSSS 194  
 Db 19 STSSSSSSSS 29

RESULT 14  
 E96685  
 probable disease resistance protein F15E12.17 [Imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: E96685  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpsiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.; Dewar, K.; Jensen, N.F.; Hughes, B.; Hultzar, L.

Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzalla Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719  
 A:Accession: E96685  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-429 <STO>  
 A:Cross-references: GB:AE005173; NID:911038485; PIDN:AAG27764.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: F15E12.17  
 A:Map position: 1

Query Match 1.8%; Score 11; DB 2; Length 429;  
 Best Local Similarity 100.0%; Pred. No. 0.16;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 184 STSSSSSSSS 194  
 Db 4 STSSSSSSSS 14

RESULT 15  
 T01552  
 hypothetical protein A\_TM018A10.2 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 19-Feb-1999 #sequence\_revision 19-Feb-1999 #text\_change 24-Mar-1999  
 C:Accession: T01552  
 R:Dempsey, S.; Harper, M.  
 A:Description: The EMBL Data Library, July 1997  
 A:Reference number: 214348  
 A:Accession: T01552  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-575 <DEM>  
 A:Cross-references: EMBL:AF013294; NID:g2252848; PID:g2252850  
 A:Experimental source: cultivar Columbia  
 C:Genetics:  
 A:Map position: 4  
 A:Introns: 271/3  
 A:Note: A\_TM018A10.2

Query Match 1.8%; Score 11; DB 2; Length 575;  
 Best Local Similarity 100.0%; Pred. No. 0.2;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 184 STSSSSSSSS 194  
 Db 59 STSSSSSSSS 69

Search completed: January 22, 2002, 16:32:17  
 Job time: 177 sec







```

RESULT 2
YC18_HUMAN STANDARD; PRT; 864 AA.
AC Q9ULX2;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN KIAA1218 (FRAGMENT).
GN KIAA1218.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K.-I., Kikuno R., Hirose M., Nomura N.,
RA Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:337-345(1999).
CC -1- SIMILARITY: BELONGS TO THE ATAXIN 7 FAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; AB033044; BAA86532.1; -.
CC KW Hypothetical protein.
CC FT NON_TER 1
CC FT DOMAIN 223 226 POLY-THR.
CC FT DOMAIN 652 664 POLY-SER.
CC FT DOMAIN 786 790 POLY-SER.
CC SQ SEQUENCE 864 AA; 92138 MW; DESCE8130E4BDA23 CRC64;

Query Match 1.9%; Score 12; DB 1; Length 864;
Best Local Similarity 100.0%; Pred. No. 0.024;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 TSSSSSSSSSS 194
Db 652 TSSSSSSSSSS 663

RESULT 3
BDNF_CAVPO STANDARD; PRT; 255 AA.
AC 070183;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE BRAIN-DERIVED NEUROTROPHIC FACTOR PRECURSOR (BDNF).
GN BDNF.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystriognathi; Cavidae; Cavia.
OX NCBI_Taxid=10141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HARTLEY WHITE; TISSUE=Liver;
RA Itoue M., Nakayama C., Noguchi H.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PROMOTES THE SURVIVAL OF NEURONAL POPULATIONS THAT ARE
CC ALL LOCATED EITHER IN THE CENTRAL NERVOUS SYSTEM OR DIRECTLY

```

```

CC CC CONNECTED TO IT (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO THE NGF-BETA FAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; AB012097; BAA25176.1; -.
CC DR Interpro; IPR002072; NGF.
CC DR Pfam; PF00243; NGF_1.
CC DR PRINTS; PD00268; NGF.
CC DR ProDom; PD002052; NGF; 1.
CC DR SMART; SM00140; NGF; 1.
CC DR PROSITE; PS00248; NGF_1; 1.
CC DR PROSITE; PS50270; NGF_2; 1.
CC KW Growth factor; Signal.
CC FT SIGNAL 1 18
CC FT PROPEL 19 136 POTENTIAL.
CC FT CHAIN 137 255 BY SIMILARITY.
CC FT DISULFID 149 216 BRAIN-DERIVED NEUROTROPHIC FACTOR.
CC FT DISULFID 194 245 BY SIMILARITY.
CC FT DISULFID 204 247 BY SIMILARITY.
CC FT CARBOHYD 129 129 N-LINKED (GLCNAC ...) (POTENTIAL).
CC SQ SEQUENCE 255 AA; 28308 MW; BA95BA3EBB8F04 CRC64;

Query Match 1.8%; Score 11; DB 1; Length 255;
Best Local Similarity 100.0%; Pred. No. 0.063;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 TSSSSSSSSSL 195
Db 57 TSSSSSSSSSL 67

RESULT 4
YG3A_YEAST STANDARD; PRT; 816 AA.
AC P53278;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHETICAL 92.7 KDA PROTEIN IN ASN2-PH1 INTERGENIC REGION.
GN YGR130C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_Taxid=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA van Dyck L., Skala J., de Wergifosse P., Purnelle B., Talia E.,
RA Nawrocki A., del Bino S., Goffeau A.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; Z72915; CA97143.1; -.
CC DR HSSP; P56255; LPJR.
CC DR SGD; S0003362; YGR130C.
CC KW Hypothetical protein.
CC FT DOMAIN 164 185 POLY-SER.
CC FT DOMAIN 205 218 POLY-THR.

```



```

OY 184 STSSSSSSSS 193
   |||||||||
DB 9 STSSSSSSSS 18

RESULT 7
ID UBC6_MOUSE STANDARD: PRT: 193 AA.
AC PS2482:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE UBIQUITIN-CONJUGATING ENZYME E2-21 KDA UBC6 (EC 6.3.2.19) (UBIQUITIN-
DE PROTEIN LIGASE) (UBIQUITIN CARRIER PROTEIN).
GN UBC6 OR UBCM3 OR UBC5.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId:10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE-Brain;
RX MEDLINE-96162026; PubMed-8576256.
RA Matuszewski K., Hauser H.P., Tietler M., Jentsch S.;
RT "Identification of a novel family of ubiquitin-conjugating enzymes
RT with distinct amino-terminal extensions."
RJ J. Biol. Chem. 271:2789-2794(1996).
RL
CC -1- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO OTHER
CC PROTEINS. MEDICATES THE SELECTIVE DEGRADATION OF SHORT-LIVED AND
CC ANOMAL. PROTEINS.
CC -1- CATALYTIC ACTIVITY: ATP + UBIQUITIN + PROTEIN LYSINE =
CC AAP + PYROPHOSPHATE + PROTEIN N-UBIQUITYLYLlysine.
CC -1- PATHWAY: SECOND STEP IN UBIQUITIN CONJUGATION.
CC -1- MISCELLANEOUS: A CYSTEINE RESIDUE IS REQUIRED FOR
CC UBIQUITIN-THIOLESTER FORMATION (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE UBIQUITIN-CONJUGATING ENZYME FAMILY.
CC SROMCER*, TO DROSOPHILA UBC22.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X92665; CA63353.1; -.
DR HSSG; P15731; 2UCE.
DR MGD; MG1:107411; Ubc65.
DR InterPro: IPR000608; UBO_conjugat.
DR Pfam; PF00179; UO_con; 1.
DR SMART; SM00212; UBCc; 1.
DR PROSITE; PS00183; UBIQUITIN_CONJUGAT_1; 1.
DR PROSITE; PS0127; UBIQUITIN_CONJUGAT_2; 1.
KW Ubiquitin conjugation; Ligase; Multigene family.
FT DOMAIN 9 Polv-SBR.
FT BINDING 131 131 UBIQUITIN (BY SIMILARITY).
FT SEQUENCE 193 AA; 21333 MW; B535B3095EFC445 CRC64;

Query Match 1.6%; Score 10; DB 1; Length 193;
Best Local Similarity 100.0%; Pred. No. 0.36;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 184 STSSSSSSSS 193
   |||||||||
DB 9 STSSSSSSSS 18

RESULT 8
ID AXIB_ARATH STANDARD: PRT: 246 AA.
AC Q38829;

```

DT 20-AUG-2001 (Rel. 40, Created)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE AUXIN-RESPONSIVE PROTEIN IAA1 (INDOLEACETIC ACID-INDUCED PROTEIN 11)  
GN IAA1 OR AT4G28640 OR T5F17.90.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_taxid=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RX MEDLINE=20083488; PubMed=106117198;  
RA Mayer K.F.X., Schnellier C., Wandt R., Murphy G., Voickert G.,  
RA Pohl T., Duesthoeft A., Stiekema W., Ertl K.-D., Terryn N.,  
RA Harris B., Ansgor W., Brand P., Griwall L.A., Rieger M.,  
RA Weichselgartner M., de Simone V., Obermair B., Mache R., Mueller M.,  
RA Kreis M., Delseny M., Puigdemont P., Watson M., Schmidt T.,  
RA Reichert B., Portelle D., Perez-Alonso M., Bouly M., Bancroft I.,  
RA Vos P., Hohselt J., Zimmermann W., Medler H., Ridley P.,  
RA Langham S.A., McCullagh B., Bilham L., Robben J.,  
RA Van der Schueren J., Grymoprez B., Chuang Y.-J., Vandenbussche F.,  
RA Braeken M., Weltjens I., Voet M., Bastiaens U., Aert R., Deoor E.,  
RA Weltekneger T., Bothé G., Ramspeniger U., Hilbert H., Braun M.,  
RA Holzner E., Brandt A., Peters S., van Staveren M., Dirks P.,  
RA Moilman P., Klein lankhorst R., Rose M., Hauf J., Koettler P.,  
RA Bernelsner S., Hempel S., Feldausch M., Lamberth S., Van den Daele H.,  
RA De Keyser A., Buysnaert C., Gielen J., Villarroel R., De Clercq R.,  
RA Van Montagu M., Rogers J., Cronin J., Quall M., Bray-Allen S.,  
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McWay K., Mayes R.,  
RA Piettek A., Rajandream M.-A., Lyne M., Benes V., Rechmann S.,  
RA Borkova D., Bloeker H., Scharfe M., Grimm M., Loeinert T.-H.,  
RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Partman B., Granderth K., Danner D., Herzl A.,  
RA Neumann S., Argitrou A., Vitale D., Liguori R., Piravandi E.,  
RA Massenet O., Quigley F., Clabaud G., Muendlein A., Felber R.,  
RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,  
RA Chedof F., Cooke R., Berger C., Montfort A., Casacuberta E.,  
RA Gibbons T., Weber N., Vandenbol M., Baryes M., Terol J., Torres A.,  
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,  
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Biele C.,  
RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,  
RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,  
RA Parnell L., Dedhia N., Gao J. L., Schütz K., Huang E., Spielgel L.,  
RA Sehn M., Murray J., Shee P., Cordes M., Abu-Threideh J.,  
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,  
RA Latrelle P., Courtney J., Cloud J., Abbott A., Scott K., Johnson D.,  
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,  
RA Kramer J., Fulton L., Wards E., Dante M., Pepin K., Hiller L.,  
RA Nelson J., Spielon J., Ryan E., Andrews S., Gelsel C., Layman D.,  
RA Du H., Ali J., Berghoff A., Jones K., Dyne K., Cotton M., Joshi C.,  
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C.,  
RA Ma P., Zhong J., Preston R., Vill D., Shekher M., Matero A., Shah R.,  
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,  
RA Granat S., Shohdy N., Hasegawa A., Hamed A., Lodhi M., Johnson A.,  
RA Chen E., Maier M., Martienssen R., McComie W.R.;  
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis  
RT thaliana.";  
RL Nature 402:769-777(1999).  
CC -I- FUNCTION: COULD ACT AS REGULATOR OF GENES RESPONSIBLE FOR  
CC MEDIATING THE VARIOUS AUXIN-INDUCED EVENTS RESPONSIBLE FOR CELL  
CC GROWTH (BY SIMILARITY).  
CC -I- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
CC -I- INDUCTION: BY AUXIN.



CC -----  
 CC -1- SIMILARITY: BELONGS TO THE AUX/IAA FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U18413; AAC9052.1; -  
 CC EMBL: AL161573; CAB81452.1; -  
 CC Mendeil: 7127; Arath:1524;7127.  
 CC Interpro: IPR003311; AUX\_IAA.  
 CC Pfam: PF02309; AUX\_IAA; 1.  
 CC Multigene family: Nuclear protein; Translation regulation.  
 CC DOMAIN 57 67  
 CC SEQUENCE 246 AA; 26516 MW; 449054E312FBDAD CRC64;  
 CC -----  
 CC Query Match 1.6%; Score 10; DB 1; Length 246;  
 CC Best Local Similarity 100.0%; Pred. No. 0.44;  
 CC Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC -----  
 CC OY 186 SSSSSSSSS 195  
 CC Db 57 SSSSSSSSS 66  
 CC -----  
 CC RESULT 9  
 CC ID LYTE\_BACSU STANDARD; PRT; 334 AA.  
 CC AC P54421;  
 CC DT 01-OCT-1996 (Rel. 34, Created)  
 CC DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 CC DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 CC DE PUTATIVE ENDOPEPTIDASE LYTE PRECURSOR (PHOSPHATASE-ASSOCIATED PROTEIN  
 CC DE PAPO) (CELL WALL-ASSOCIATED POLYPEPTIDE CMBP33).  
 CC OS LYTE OR PAPO.  
 CC GN Bacillus subtilis.  
 CC OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 CC OC Bacillus/Staphylococcus group; Bacillus.  
 CC OX NCBI\_TaxID=1423;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A., AND SEQUENCE OF 26-36.  
 CC RC STRAIN=168;  
 CC RX MEDLINE=98117063; PubMed=9457885;  
 CC RA Margot P., Mahlen M., Gholamshahian A., Piggot P., Karamata D.;  
 CC RT "The lyte gene of Bacillus subtilis 168 encodes a cell wall  
 CC RT hydrolase";  
 CC RL J. Bacteriol. 180:749-752(1998).  
 CC RN [2]  
 CC RP SEQUENCE OF 311-334 FROM N.A.  
 CC RC STRAIN=168 / SMY;  
 CC RX MEDLINE=94321340; PubMed=8045898;  
 CC RA Jin S., Sonenshein A.L.;  
 CC RT "Identification of two distinct Bacillus subtilis citrate synthase  
 CC RT genes";  
 CC RL J. Bacteriol. 176:4669-4679(1994).  
 CC CC -1- FUNCTION: PUTATIVE D-GLUTAMATE - M-DIAMINOPIMLATE ENDOPEPTIDASE.  
 CC CELL WALL HYDROLASE INVOLVED IN CELL AUTOLYSIS.  
 CC -1- SUBCELLULAR LOCATION: CELL-WALL BOUND.  
 CC -1- DOMAIN: LYSM REPEATS ARE THOUGHT TO BE INVOLVED IN PEPTIDOLYCAN  
 CC BINDING.  
 CC -1- SIMILARITY: BELONGS TO THE E-COLI NLPC / LISTERIA P60 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

CC -----  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U38819; AAC25975.1; -  
 CC EMBL: U05256; -; NOT\_ANNOTATED\_CDS.  
 CC Subtilist: BG11406; LYTE.  
 CC Interpro: IPR002482; LYSM.  
 CC Interpro: IPR000064; NLPC\_P60.  
 CC Pfam: PF01476; LYSM; 3.  
 CC Pfam: PF00877; NLPC\_P60; 1.  
 CC SMART: SM00257; LYSM; 3.  
 CC Cell wall; Hydrolase; Repeat; Signal; Complete proteome.  
 CC KW SIGNAL 1 25  
 CC FT CHAIN 26 334  
 CC FT REPEAT 28 70  
 CC FT REPEAT 88 130  
 CC FT REPEAT 151 193  
 CC FT DOMAIN 73 86  
 CC FT DOMAIN 132 141  
 CC FT DOMAIN 144 149  
 CC FT DOMAIN 204 211  
 CC SEQUENCE 334 AA; 35455 MW; D809C7DD4BAC475D CRC64;  
 CC -----  
 CC Query Match 1.6%; Score 10; DB 1; Length 334;  
 CC Best Local Similarity 100.0%; Pred. No. 0.57;  
 CC Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC -----  
 CC OY 184 STSSSSSSSS 193  
 CC Db 132 STSSSSSSSS 141  
 CC -----  
 CC RESULT 10  
 CC ID PL5B\_PHAVU STANDARD; PRT; 461 AA.  
 CC AC 043822;  
 CC DT 15-DEC-1998 (Rel. 37, Created)  
 CC DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 CC DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 CC DE GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE PRECURSOR (EC 2.3.1.15) (GPAT).  
 CC GN PL5B.  
 CC OS Phaseolus vulgaris (kidney bean) (French bean).  
 CC OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 CC OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.  
 CC OX NCBI\_TaxID=3885;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=CV. ANNABEL; TISSUE=leaf;  
 CC RX MEDLINE=95232196; PubMed=7716242;  
 CC RA Fritz M., Heinz E., Wolter F.P.;  
 CC RT "Cloning and sequencing of a full-length cDNA coding for  
 CC RT sn-glycerol-3-phosphate acyltransferase from Phaseolus vulgaris";  
 CC RL Plant Physiol. 107:1039-1040(1995).  
 CC CC -1- FUNCTION: ESTERIFIES ACYL-GROUP FROM ACYL-ACP TO THE SN-1 POSITION  
 CC OF GLYCEROL-3-PHOSPHATE. THE ENZYME FROM CHILLING-RESISTANT PLANTS  
 CC DISCRIMINATE AGAINST NON-FLUID PALMITIC ACID AND SELECT OLEIC ACID  
 CC WHEREAS THE ENZYME FROM SENSITIVE PLANTS ACCEPTS BOTH FATTY ACIDS.  
 CC -1- CATALYTIC ACTIVITY: ACYL-COA + SN-GLYCEROL 3-PHOSPHATE =  
 CC COA + 1-ACYL-SN-GLYCEROL 3-PHOSPHATE.  
 CC -1- PATHWAY: FIRST STEP IN DE NOVO PHOSPHOLIPID BIOSYNTHESIS. IT MAY  
 CC ALSO FUNCTION IN THE REGULATION OF MEMBRANE BIOGENESIS.  
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST STROMA.  
 CC -1- SIMILARITY: BELONGS TO THE PLANT GPAT FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: X79722; CAA56159.1; -  
 DR InterPro: IPR002123; Acyltransferase.  
 KW Phospholipid biosynthesis; Transferase; Acyltransferase;  
 FT TRANSIT 1 96 CHLOROPLAST (POTENTIAL).  
 FT CHAIN 97 461 GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE.  
 SO SEQUENCE 461 AA; 50697 MW; EAC4FC837908B38A CRC64;

Query Match 1.6%; Score 10; DB 1; Length 461;  
 Best Local Similarity 100.0%; Pred. No. 0.74;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 183 TSSSSSSSS 192  
 DB 49 SSTSSSSSS 58

RESULT 11  
 ID KC3H\_DICDI STANDARD; PRT; 468 AA.  
 AC P51136;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE GLYCOCEN SYNTHASE KINASE-3 HOMOLOG (EC 2.7.1.1) (GSK-3).  
 GN GSKA.  
 OS Dictyostelium discoideum (Slime mold).  
 OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.  
 OX NCBI\_TaxID=44689;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AX2;  
 RX MEDLINE=95112340; PubMed=7813009;  
 RA Hatwood A.J., Plyte S.E., Woodgett J., Strutt H., Kay R.R.;  
 RL "Glycogen synthase kinase 3 regulates cell fate in Dictyostelium";  
 Cell 80:139-148(1995).  
 CC -1- FUNCTION: DURING CELLULAR DIFFERENTIATION IT MAY MEDIATE A CAMP  
 STIMULATED SIGNAL TRANSDUCTION PATHWAY THAT REGULATES PRESORE AND  
 PRESTALK B CELL PROPORTIONS.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC CDC2/CDKX SUBFAMILY; GSK-3 SUBSUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: L34674; AAA65968.1; -  
 DR HSSP: P24941; IAO1.  
 DR Dictydb; DD01073; gskA.  
 DR InterPro; IPR000719; Euk\_kinase.  
 DR InterPro; IPR002290; Ser\_thr\_kin\_actsite.  
 DR Pfam; PF00069; kinase; 1.  
 DR SMART; SMO0220; S\_TKC; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW transferase; Serine/threonine-protein kinase; ATP-binding;  
 KW phosphorylation.  
 FT DOMAIN 57 340 PROTEIN KINASE.  
 FT NP\_BIND 63 71 ATP (BY SIMILARITY).  
 FT BINDING 86 86 ATP (BY SIMILARITY).  
 FT ACT\_SITE 180 180 BY SIMILARITY.  
 FT MOD\_RES 215 215 PHOSPHORYLATION (BY SIMILARITY).  
 FT DOMAIN 23 28 POLY-SER.  
 FT DOMAIN 29 38 POLY-SER.  
 FT DOMAIN 405 410 POLY-SER.  
 FT DOMAIN 429 433 POLY-SER.  
 FT DOMAIN 456 465 POLY-THR.

SO SEQUENCE 468 AA; 51598 MW; 50DCE4B7D601D5F9 CRC64;

Query Match 1.6%; Score 10; DB 1; Length 468;  
 Best Local Similarity 100.0%; Pred. No. 0.75;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 185 TSSSSSSSS 194  
 DB 28 TSSSSSSSS 37

RESULT 12  
 ID YOD0\_YEAST STANDARD; PRT; 484 AA.  
 AC 008193;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE HYPOTHETICAL 51.9 KDA PROTEIN IN MSEL-LAG2 INTERGENIC REGION  
 DE PRECURSOR.  
 GN YOL030W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Habbig B., Hattenhorst U., Hollenberg C.P., Ramezani Rad M.;  
 RL Submitted (JUL-1996) to the EMBL/Genbank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE GAS1 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z74772; CAA99030.1; -  
 DR SGD: S0005390; YOL030W.  
 DR Hypothetical protein; Glycoprotein; Signal.  
 FT SIGNAL 1 22 POTENTIAL.  
 FT CHAIN 23 484 HYPOTHETICAL PROTEIN YOL030W.  
 FT DOMAIN 395 403 ASP/GLU-RICH (HIGHLY ACIDIC).  
 FT DOMAIN 405 460 SER-RICH.  
 FT DOMAIN 405 427 POLY-SER.  
 FT DOMAIN 432 439 POLY-SER.  
 FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 60 60 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 344 344 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 359 359 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SO SEQUENCE 484 AA; 51869 MW; E65B9F4689BBB34C CRC64;

Query Match 1.6%; Score 10; DB 1; Length 484;  
 Best Local Similarity 100.0%; Pred. No. 0.77;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 185 TSSSSSSSS 194  
 DB 404 TSSSSSSSS 413

RESULT 13  
 ID Y859\_TREPA STANDARD; PRT; 494 AA.  
 AC 083831; O83832;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)

```
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN TP0859/TP0860 PRECURSOR.
GN TP0859/TP0860.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOLS;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwin M., Jockey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utechtack T.,
RA McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete."
RL Science 281:375-388(1998).
CC -1- SIMILARITY: BELONGS TO THE UPF0164 FAMILY.
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION. REF.1 SEQUENCE DIFFERS
CC FROM THAT SHOWN DUE TO A FRAMESHIFT IN POSITION 232.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001256; AAC65831.1; ALT_FRAME.
DR EMBL; AE001256; AAC65832.1; ALT_FRAME.
DR TIGR; TP0859; -.
DR TIGR; TP0860; -.
KW Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 1 44 POTENTIAL.
FT CHAIN 45 494 HYPOTHETICAL PROTEIN TP0859/TP0860.
FT SEQUENCE 494 AA; 52883 MW; 5B419C60CEAF840 CRC64;
SQ
Query Match 1.6%; Score 10; DB 1; Length 494;
Best Local Similarity 100.0%; Pred. No. 0.79;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 8 AGSGGGGGG 17
DB 296 AGSGGGGGG 305
RESULT 14
YN23_YEAST STANDARD; PRT; 503 AA.
AC P53832;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 52.3 KDA PROTEIN IN MRPL10-ERG24 INTERGENIC REGION
DE PRECURSOR.
GN YNL283C OR N0583.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Mesenney F., Dubois E., Vierendeels F., Scherens B., Pierard A.,
RA Glansdorff N.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO S.POMBE SPBC3D5.14C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z71559; CAA96195.1; -.
DR SGD; S0005227; YNL283C.
DR InterPro; IPR002889; MSC.
DR Pfam; PF01822; MSC; 1.
DR SMART; SM00321; MSC; 1.
KW Hypothetical protein; Transmembrane; Signal.
FT SIGNAL 1 23
FT CHAIN 24 503 HYPOTHETICAL PROTEIN YNL283C.
FT TRANSMM 61 81 POTENTIAL.
FT TRANSMEM 326 346 POTENTIAL.
FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 429 429 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 475 475 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SEQUENCE 503 AA; 52292 MW; F2392A73C5CBAB50 CRC64;
SQ
Query Match 1.6%; Score 10; DB 1; Length 503;
Best Local Similarity 100.0%; Pred. No. 0.8;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 185 TSSSSSSSS 194
DB 182 TSSSSSSSS 191
RESULT 15
MS2P_HUMAN STANDARD; PRT; 519 AA.
ID MS2P_HUMAN
AC O43462;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE STEROL-REGULATORY ELEMENT-BINDING PROTEIN INTRAMEMBRANE PROTEASE
DE (EC 3.4.24.-) (SITE-2 PROTEASE).
GN S2P.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fibroblast;
RX MEDLINE=98324087; PubMed=9659902;
RA Rawson R.B., Zelenski N.G., Nijhawan D., Ye J., Sakai J., Hasan M.T.,
RA Chang T.Y., Brown M.S., Goldstein J.L.;
RT "Complementation cloning of S2P, a gene encoding a putative
RT metalloprotease required for intramembrane cleavage of SREBPs."
RL Mol. Cell 1:47-57(1997).
CC -1- FUNCTION: INTRAMEMBRANE PROTEOLYSIS OF STEROL-REGULATORY ELEMENT-
CC BINDING PROTEINS (SREBPS) WITHIN THE FIRST TRANSMEMBRANE SEGMENT
CC THEREBY RELEASING THE N-TERMINAL SEGMENT WITH A PORTION OF THE
CC TRANSMEMBRANE SEGMENT ATTACHED. SITE-2 CLEAVAGE COMES AFTER SITE-1
CC CLEAVAGE WHICH TAKE PLACES IN THE LOWER LOOP.
CC -1- COFACTOR: BINDS ONE ZINC ION (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN HEART, BRAIN, PLACENTA, LUNG,
CC LIVER, MUSCLE, KIDNEY AND PANCREAS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M50 (ZINC
CC METALLOPROTEASE).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
DR EMBL; AF019612; AAC51937.1; -.  
DR MEROPS; M50.001; -.  
DR InterPro; IPR001193; Peptidase\_M50.  
DR InterPro; IPR000130; Zn\_MTPeptidse.  
DR Pfam; PF02163; Peptidase\_M50; 1.  
DR PRINTS; PR01000; SRBPS2PTASE.  
DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
KW Hydrolyase; Metalloprotease; Zinc; Transmembrane.  
FT METAL 171 171 ZINC (CATALYTIC) (BY SIMILARITY).  
FT ACT\_SITE 172 172 BY SIMILARITY.  
FT METAL 175 175 ZINC (CATALYTIC) (BY SIMILARITY).  
FT TRANSMEM 4 24 POTENTIAL.  
FT TRANSMEM 75 95 POTENTIAL.  
FT TRANSMEM 186 206 POTENTIAL.  
FT TRANSMEM 229 249 POTENTIAL.  
FT TRANSMEM 448 468 POTENTIAL.  
FT TRANSMEM 493 513 POTENTIAL.  
FT DOMAIN 109 136 POLY-SER.  
FT DOMAIN 380 384 POLY-SER.  
FT MUTAGEN 171 171 H->F: ACTIVITY ABOLISHED.  
FT MUTAGEN 172 172 E->A: ACTIVITY ABOLISHED.  
FT MUTAGEN 172 172 E->O: ACTIVITY ABOLISHED.  
FT MUTAGEN 172 172 E->D: ACTIVITY PARTIALLY ABOLISHED.  
FT MUTAGEN 175 175 H->F: ACTIVITY ABOLISHED.  
SQ SEQUENCE 519 AA; 57443 MW; 247D69EDFD7747BD CRC64;

Query Match 1.6%; Score 10; DB 1; Length 519;  
Best Local Similarity 100.0%; Pred. No. 0.82;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 186 SSSSSSSSSL 195  
|||  
Db 128 SSSSSSSSSL 137

Search completed: January 22, 2002, 16:37:08  
Job time: 318 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2002, 16:31:00 ; Search time 40.87 Seconds  
(without alignments)  
2229.096 Million cell updates/sec

Title: US-09-485-529-7

Perfect score: 623  
Sequence: 1 MKREYDAGSGGGGSGGMS.....TIGMHTPLTASAMRLAGP 623

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 473505 seqs, 146272329 residues

Word size : 0

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

- 1: SPTEMBL\_17:\*
- 2: sp.archaea:\*
- 3: sp.bacteria:\*
- 4: sp.fungi:\*
- 5: sp.human:\*
- 6: sp.invertebrate:\*
- 7: sp.mammal:\*
- 8: sp.mbc:\*
- 9: sp.organelle:\*
- 10: sp.phage:\*
- 11: sp.plant:\*
- 12: sp.potent:\*
- 13: sp.virus:\*
- 14: sp.yeast:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	623	100.0	623	10	Q9ST59
2	132	21.2	620	10	Q9ST48
3	89	14.3	625	10	Q9MB96
4	23	3.7	547	10	Q9SRP9
5	22	3.5	511	10	Q9C8Y3
6	22	3.5	662	10	Q65367
7	19	3.0	493	10	Q9AS97
8	19	3.0	523	10	Q9LF53
9	19	3.0	532	10	Q23643
10	19	3.0	533	10	Q23724
11	19	3.0	533	10	Q9LQ78
12	19	3.0	587	10	Q23725
13	19	3.0	587	10	Q23642
14	19	3.0	587	10	Q23643
15	16	2.6	275	10	Q9SLH3
16	12	1.9	118	5	Q9U1K5
17	12	1.9	119	5	Q9V401
18	12	1.9	153	3	Q08906
19	12	1.9	229	11	Q9JM93

20	12	1.9	390	2	Q9PHR0
21	12	1.9	409	3	Q9HED2
22	12	1.9	446	3	Q9Y849
23	12	1.9	574	4	Q9BT08
24	12	1.9	856	5	Q9GZ10
25	12	1.9	989	5	Q9W2S4
26	12	1.9	1297	4	Q15025
27	12	1.9	1383	12	Q89232
28	12	1.9	2161	12	Q9E1P1
29	12	1.9	2164	12	Q9DH52
30	12	1.8	164	5	Q19817
31	12	1.8	187	11	Q99NV0
32	12	1.8	193	11	Q99NV1
33	12	1.8	216	10	Q9C9J8
34	12	1.8	238	5	Q44958
35	12	1.8	291	10	Q9M160
36	12	1.8	318	10	Q9S782
37	12	1.8	321	5	Q9W3Z5
38	12	1.8	371	10	Q9FN34
39	12	1.8	417	5	Q9XY07
40	12	1.8	429	10	Q9C515
41	12	1.8	450	11	Q9DA19
42	12	1.8	575	10	Q23083
43	12	1.8	600	13	Q90237
44	12	1.8	614	5	P91720
45	12	1.8	632	10	Q80569

#### ALIGNMENTS

RESULT	ID	Query Match	Score	DB	Length	Matches	Conservative	Mismatches	Indels	Gaps
1	Q9ST59	100.0%	623	10	623	623	0	0	0	0
2	Q9ST59	100.0%	623	10	623	623	0	0	0	0
3	Q9ST59	100.0%	623	10	623	623	0	0	0	0
4	Q9ST59	100.0%	623	10	623	623	0	0	0	0
5	Q9ST59	100.0%	623	10	623	623	0	0	0	0
6	Q9ST59	100.0%	623	10	623	623	0	0	0	0
7	Q9ST59	100.0%	623	10	623	623	0	0	0	0
8	Q9ST59	100.0%	623	10	623	623	0	0	0	0
9	Q9ST59	100.0%	623	10	623	623	0	0	0	0
10	Q9ST59	100.0%	623	10	623	623	0	0	0	0
11	Q9ST59	100.0%	623	10	623	623	0	0	0	0
12	Q9ST59	100.0%	623	10	623	623	0	0	0	0
13	Q9ST59	100.0%	623	10	623	623	0	0	0	0
14	Q9ST59	100.0%	623	10	623	623	0	0	0	0
15	Q9ST59	100.0%	623	10	623	623	0	0	0	0
16	Q9ST59	100.0%	623	10	623	623	0	0	0	0
17	Q9ST59	100.0%	623	10	623	623	0	0	0	0
18	Q9ST59	100.0%	623	10	623	623	0	0	0	0
19	Q9ST59	100.0%	623	10	623	623	0	0	0	0
20	Q9ST59	100.0%	623	10	623	623	0	0	0	0
21	Q9ST59	100.0%	623	10	623	623	0	0	0	0
22	Q9ST59	100.0%	623	10	623	623	0	0	0	0
23	Q9ST59	100.0%	623	10	623	623	0	0	0	0
24	Q9ST59	100.0%	623	10	623	623	0	0	0	0
25	Q9ST59	100.0%	623	10	623	623	0	0	0	0
26	Q9ST59	100.0%	623	10	623	623	0	0	0	0
27	Q9ST59	100.0%	623	10	623	623	0	0	0	0
28	Q9ST59	100.0%	623	10	623	623	0	0	0	0
29	Q9ST59	100.0%	623	10	623	623	0	0	0	0
30	Q9ST59	100.0%	623	10	623	623	0	0	0	0
31	Q9ST59	100.0%	623	10	623	623	0	0	0	0
32	Q9ST59	100.0%	623	10	623	623	0	0	0	0
33	Q9ST59	100.0%	623	10	623	623	0	0	0	0
34	Q9ST59	100.0%	623	10	623	623	0	0	0	0
35	Q9ST59	100.0%	623	10	623	623	0	0	0	0
36	Q9ST59	100.0%	623	10	623	623	0	0	0	0
37	Q9ST59	100.0%	623	10	623	623	0	0	0	0
38	Q9ST59	100.0%	623	10	623	623	0	0	0	0
39	Q9ST59	100.0%	623	10	623	623	0	0	0	0
40	Q9ST59	100.0%	623	10	623	623	0	0	0	0
41	Q9ST59	100.0%	623	10	623	623	0	0	0	0
42	Q9ST59	100.0%	623	10	623	623	0	0	0	0
43	Q9ST59	100.0%	623	10	623	623	0	0	0	0
44	Q9ST59	100.0%	623	10	623	623	0	0	0	0
45	Q9ST59	100.0%	623	10	623	623	0	0	0	0

OY	181	GGTSSSSSSSSSLGGGASSVVEAAPPAAANATPALPVVVDTQEGITVLHALLA	240
Db	181	GGSTSSSSSSSSSLGGGARSVVEAAPVAAAANNATPALPVVVDTQEGILVALHALLA	240
OY	241	CAEAVOENTLSAEALVKQIPLLASOGAMRKVYAFCGEALARVFREFRPQDSSLDA	300
Db	241	CAEAVOENTLSAEALVKQIPLLAASGGAMRKVYAFCGELARVRFRRPQDSSLDA	300
OY	301	AFAOLLNAHFESCPYLKFAHTANAQLAEAFAGCRHVHYVDGIKQMOMPALLOLAL	360
Db	301	AFAOLLNAHFESCPYLKFAHTANOQLAEAFAGCRHVHYVDGIKQMOMPALLOLAL	360
OY	361	RPGPPSRFLGVGPPOPDFETDALQQYGMLAQPAHTIRVDFQYRGIVAAATLDLEPPML	420
Db	361	RPGPPSRFLGVGPPOPDFETDALQQYGKMLAQPAHTIRVDFQYRGLVAATLDLEPPML	420
OY	421	QPGEEDPNEEPEVIAYNSVFEHMRLLAQBGALEKYLGTVRAVRPRIVTVEQEANHNNG	480
Db	421	QPGEEDPNEEPEVIAYNSVFEHMRLLAQBGALEKYLGTVRAVRPRIVTVEQEANHNNSG	480
OY	481	TFLDRFTESLHYISTMDDSLGSGSGGGSPBEVSSSGAAAAAAGTQVMSEYVLGNQICN	540
Db	481	TFLDRFTESLHYISTMDDSLGSGSGGGSPBEVSSSGAAAAAAGTQVMSEYVLGNQICN	540
OY	541	VVACEGAEERTERHETHLLQOMNRNLGNAGFETVHLGHSNAYKQASTLLALFAGDGDKYEKEE	600
Db	541	VVACEGAEERTERHETHLLQOMNRNLGNAGFETVHLGHSNNYKQASTLLALFAGDGDKYEKEE	600
OY	601	GCLTLGMHTRPLITSAMRIAGP 623	
Db	601	GCLTLGMHTRPLITSAMRIAGP 623	

RESULT	2		
09ST48			
ID	Q9ST48	PRELIMINARY;	PRT; 630 AA.
AC	Q9ST48;		
DT	01-MAY-2000 (TEMBREL. 13, Created)		
DT	01-MAY-2000 (TEMBREL. 13, Last sequence update)		
DT	01-MAR-2001 (TEMBREL. 16, Last annotation update)		
DE	GIBBERELLIN RESPONSE MODULATOR (FRAGMENT).		
DN	DB.		
OS	zma_mays (Maize).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade		
OC	Panicoidae; Andropogoneae; Zea.		
OX	NCBI_TaxID=4577;		
XP	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=99347734; PubMed=10421366;		
RA	Peng J., Richards D.E., Hartley N.M., Murphy G.P., Devos K.M.,		
RA	Flintham J.E., Beales J., Fish L.U.J., Worland A.J., Pelica F.,		
RA	Sudhakar D., Christou P., Snape J.W., Gale M.D., Harberd N.P.;		
RT	"Green revolution" genes encode mutant gibberellin response		
RL	modulators.";		
RL	Nature 400:256-261(1999).		
DR	EMBL, AJ242530; CAB51557.1; .		
FT	NON_TER 630		
QZ	SEQUENCE 630 AA; 66028 MW; 3D56851726C51042 CRC64;		

	Query Match	21.2%	Score 132	DB 10	Length 630
	Best Local Similarity	100.0%	Pred. No. 1.8e-115		
	Matches 132	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	293	PDSLLDAAFADLHAHNYESCPLYLKFAHFTANQAILLEAFACGRVHVYDFEIKGMQMP	352		
Db	302	PDSLLDAFAFDLHAHNYESCPLYLKFAHFTANQAILLEAFACGRVHVYDFEIKGMQMP	361		
QY	353	ALLQALALRPGSPSFRITLGVGPPQPDETDALQQYQGWKLAQPAHNTITVDYQRLGVAATL	412		
Db	362	ALLQALALRPGSPSFRITLGVGPPQPDETDALQQYQGWKLAQPAHNTITVDYQRLGVAATL	421		

```
QY      413 ADLEPFMLQPEG 424
        |||||
Db      422 ADLEPFMLQPEG 433
```

	RESULT	3
Q9MB96		
ID	Q9MB96	PRELIMINARY; PRT; 625 AA.
AC	Q9MB96	
DT	01-OCT-2000 (TREMBlrel. 15, Created)	
DT	01-OCT-2000 (TREMBlrel. 15, Last sequence update)	
DT	01-OCT-2000 (TREMBlrel. 15, Last annotation update)	
DE	OSGAI.	
GN	OSGAI.	
OS	Oryza sativa (Rice).	
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;	
OC	Erihatoideae; Oryzaceae; Oryza.	
OX	NCBI_TaxID=4530;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=CV. NIPPONBARE;	
RX	MEDLINE=20179680; PubMed=10713441;	
RA	Ogawa M., Kusano T., Katsumi M., Sano H.;	
RT	"Rice gibberellin-insensitive gene homolog, OSGAI, encodes a nuclear-	
RT	localized protein capable of gene activation at transcriptional	
RT	level.";	
RL	Gene 245:21-29(2000).	
DR	EMBL; AB030956; BAA90749.1; -	
SO	SEQUENCE 625 AA; 65406 MW; 034FF02719D42E97 CRC64;	

	Query Match	14.3%	Score 89;	DB 10;	Length 625;
	Best Local Similarity	100.0%	Pred. No. 5.6e-75;		
Matches	89;	Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
Qy	337	RHHVHDEGKGGOMRPA	LLQALALAPRGGRSPRLTGVGPDPDENDALDQGVWKLQAFH	396	
Db	343	RHHVHDEGKGGOMRPA	LLQALALAPRGGRSPRLTGVGPDPDENDALDQGVWKLQAFH	402	
Qy	397	TIRVDFQYRGILVAAT	LADLEPRLDPEGE	425	
Db	403	TIRVDFQYRGILVAAT	LADLEPRLDPEGE	431	

RESULT	4
Q9SRP9	PRELIMINARY; PRT; 547 AA.
Q9SRP9	
AC	Q9SRP9;
DT	01-MAY-2000 (TrEMBLrel. 13, Created)
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT	01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE	RGAI-LIKE PROTEIN.
GN	T21P5.13.
OS	Arabidopsis thaliana (Mouse-ear cress).
OC	Euraryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC	eucotsids II; Brassicales; Brassicaceae; Arabidopsis.
OX	NCBI_TaxId=3702;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=CV. COLUMBIA;
RA	Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
RA	Ronning C.M., Koo H., Fujii C.Y., Uteback T.R., Barnstead M.E.,
RA	Bowman C.L., White O., Niernan W.C., Fraser C.M.;
RT	"Arabidopsis thaliana chromosome III BAC T21P5 genomic sequence.";
RL	Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RS	EMBL; AC009895; AAF01590.1; -
SD	SEQUENCE 547 AA; 60493 MW; CAD18D5951D95634 CRC64;

Query Match 3.78; Score 23; DB 10; Length 547;

Best Local Similarity 100.0%; Pred. No. 7.3e-13;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 309 HFESECPYLKFAHFTANQALILE 331

Db 257 HFESECPYLKFAHFTANQALILE 279

RESULT 5  
Q9C8Y3 PRELIMINARY; PRT; 511 AA.

AC Q9C8Y3; 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DE GIBBERELIN REGULATORY PROTEIN, PUTATIVE.

GN T27F4.10.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

NCBI\_TaxID=3702;

SEQUENCE FROM N.A.

RC STRAIN=CV. COL-0;

RA Sanchez-Fernandez R., Ardiles-Diaz W., van Montagu M., Inze D.,

RA May M.J.;

RT "Cloning of the Arabidopsis thaliana RGA-like gene, a putative member

of the VRII domain transcription factor family";

RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL: A1224957; CAA12242.1; -

DR Mendel; 29006; Arabid; 3051; 29006.

DR InterPro; IPR001680; WD40.

DR Pfam; PF00400; WD40; 1.

DR SMART; SM00320; WD40; 1.

DR PROSITE; PS50082; WD\_REPEATS\_2; 1.

DR PROSITE; PS50294; WD\_REPEATS\_REGION; 1.

KW Repeat; WD repeat.

SQ SEQUENCE 662 AA; 73126 MW; 958AB73D42121CA8 CRC64;

Query Match 3.5%; Score 22; DB 10; Length 662;

Best Local Similarity 100.0%; Pred. No. 7.6e-12; Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 309 HFESECPYLKFAHFTANQALILE 330

Db 376 HFESECPYLKFAHFTANQALILE 397

RESULT 7  
Q9AS97 PRELIMINARY; PRT; 493 AA.

AC Q9AS97; 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DE PUTATIVE OSCAI.

GN P0707D10.30.

OS Oryza sativa (Rice).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC Ehrhartoideae; Oryzaceae; Oryza.

NCBI\_TaxID=4530;

NCBI\_TaxID=4530;

NCBI\_TaxID=4530;

NCBI\_TaxID=4530;

NCBI\_TaxID=4530;

NCBI\_TaxID=4530;

NCBI\_TaxID=4530;

NCBI\_TaxID=4530;

NCBI\_TaxID=4530;

NCBI\_TaxID=4530;

NCBI\_TaxID=4530;

NCBI\_TaxID=4530;

NCBI\_TaxID=4530;

NCBI\_TaxID=4530;

NCBI\_TaxID=4530;

NCBI\_TaxID=4530;

NCBI\_TaxID=4530;

NCBI\_TaxID=4530;

NCBI\_TaxID=4530;

NCBI\_TaxID=4530;

NCBI\_TaxID=4530;

NCBI\_TaxID=4530;

NCBI\_TaxID=4530;

NCBI\_TaxID=4530;

NCBI\_TaxID=4530;

NCBI\_TaxID=4530;

NCBI\_TaxID=4530;

NCBI\_TaxID=4530;

NCBI\_TaxID=4530;

NCBI\_TaxID=4530;

NCBI\_TaxID=4530;

NCBI\_TaxID=4530;

NCBI\_TaxID=4530;

NCBI\_TaxID=4530;

Query Match 3.0%; Score 19; DB 10; Length 493;

Best Local Similarity 100.0%; Pred. No. 3.9e-09; Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 314 CPYLKFAHFTANQALILEAF 332

Db 163 CPYLKFAHFTANQALILEAF 181

RESULT 8  
Q9LF53 PRELIMINARY; PRT; 523 AA.

AC Q9LF53; 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DE RGA-LIKE PROTEIN.

GN Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Kotani H.,  
RA Tabata S., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.,  
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Arabidopsis sequencing project;  
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL391150; CAC01893.1; -  
SQ SEQUENCE 523 AA; 57326 MW; 0F6CE0BD13403C35 CRC64;

Query Match 3.0%; Score 19; DB 10; Length 532;  
Best Local Similarity 100.0%; Pred. No. 4.1e-09;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 313 SCPYLKFAHFTANQAILLEA 331  
DB 234 SCPYLKFAHFTANQAILLEA 252

RESULT 9  
ID 023643 PRELIMINARY; PRT; 532 AA.  
AC 023643;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
DE RGA2 PROTEIN.  
GN RGA2.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=97379310; PubMed=9237632;  
RA Truong H.N., Caboche M., Daniel-Vedele F.;  
RA "Sequence and characterization of two Arabidopsis thaliana CDNAs  
RT isolated by functional complementation of a yeast gln1 mutant.";  
RL FEBS Lett. 410:213-218(1997).  
DR EMBL; Y11337; CAA72178.1; -  
DR Mendel; 24146; Arath;3051;24146.  
SQ SEQUENCE 532 AA; 58789 MW; 17AC719CA8072239 CRC64;

Query Match 3.0%; Score 19; DB 10; Length 532;  
Best Local Similarity 100.0%; Pred. No. 4.2e-09;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 CPYLKFAHFTANQAILLEAF 332  
DB 246 CPYLKFAHFTANQAILLEAF 264

RESULT 10  
ID 023724 PRELIMINARY; PRT; 532 AA.  
AC 023724;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
DE GAI PROTEIN.  
GN GAI.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LANDSBERG ERECTA;  
RX MEDLINE=98051192; PubMed=9389651;  
RA Peng J., Carol P., Richards D.E., King K.E., Cowling R.J.,  
RA Murphy G.P., Harberd N.P.;  
RT "The Arabidopsis GAI gene defines a signaling pathway that negatively  
RT regulates gibberellin responses.";  
RL Genes Dev. 11:3194-3205(1997).  
DR EMBL; Y15193; CAA5492.1; -  
DR Mendel; 24070; Arath;3051;24070.  
SQ SEQUENCE 532 AA; 58789 MW; 4AF4BC6EC4265503 CRC64;

Query Match 3.0%; Score 19; DB 10; Length 532;  
Best Local Similarity 100.0%; Pred. No. 4.2e-09;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 CPYLKFAHFTANQAILLEAF 332  
DB 246 CPYLKFAHFTANQAILLEAF 264

RESULT 11  
ID 091Q78 PRELIMINARY; PRT; 533 AA.  
AC 091Q78;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE F10B6.34.  
GN Arabidopsis thaliana (Mouse-ear cress).  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Chao Q., Shinn P., Dunn P., Buehler E., Kahn S., Kim C., Walker M.,  
RA Williams S., Altafi H., Araujo R., Conn L., Conway A.B., Gonzalez A.,  
RA Hansen N.F., Hulzar L., Kremetska I., Lenz C., Li J., Liu S.,  
RA Lueros S., Rowley D., Schwartz J., Toriumi M., Vysotskaya V., Yu G.,  
RA Davis R.W., Federspiel N.A., Theologis A., Ecker J.R.;  
RT "Genomic sequence for Arabidopsis thaliana BAC F10B6 from chromosome  
RT 1.";  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Ecker J.R.;  
RN Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Ecker J.R.;  
RN Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,  
RA Khan S., Kim C., Altafi H., Bel B., Chin C., Chou J., Choi E.,  
RA Conn L., Conway A., Gonzalez A., Hansen N., Howling B., Koo T., Lam B.,  
RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Muharsy N.,  
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,  
RA Theologis A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC006917; AAF79228.1; -  
SQ SEQUENCE 533 AA; 58926 MW; 6464B8C129D79528 CRC64;

Query Match 3.0%; Score 19; DB 10; Length 533;  
Best Local Similarity 100.0%; Pred. No. 4.2e-09;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 CPYLKFAHFTANQAILLEAF 332



Db 247 CPYLKFAHFTANQALILEAF 265

RESULT 12  
ID 023725 PRELIMINARY; PRT: 587 AA.

DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
DE GRS PROTEIN.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
NX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RA Peng J., Carol P., Richards D.E., King K.E., Cowling R.J.,  
RA Murphy G.P., Harberd N.P.;  
RA Genes Dev. 0:0-0(0).  
RL EMBL: Y15194; CAA75493.1;  
DR Mendel: 24071; Atarch:3051;24071.  
SQ SEQUENCE 587 AA; 64006 MW; F6F6C738E7DCA9 CRC64;

Query Match  
Best Local Similarity 3.0%; Score 19; DB 10; Length 587;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 CPYLKFAHFTANQALILEAF 332  
Db 299 CPYLKFAHFTANQALILEAF 317

RESULT 13  
ID 023642 PRELIMINARY; PRT: 587 AA.

DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
DE RGA1 PROTEIN.  
GN RGA1.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
NX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC MEDLINE=97379310; PubMed=9237632;  
RA Truong H.N., Caboche M., Daniel-Vedele F.;  
RT "Sequence and characterization of two Arabidopsis thaliana cDNAs  
RT isolated by functional complementation of a yeast gln3 gdh1 mutant.";  
RL FEBS Lett. 410:213-218(1997).  
DR EMBL: Y11336; CAA72177.1;  
DR Mendel: 24145; Atarch:3051;24145.  
SQ SEQUENCE 587 AA; 64023 MW; DDA7A3C741FB51EF CRC64;

Query Match  
Best Local Similarity 3.0%; Score 19; DB 10; Length 587;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 CPYLKFAHFTANQALILEAF 332  
Db 299 CPYLKFAHFTANQALILEAF 317

RESULT 14

09SLH3  
ID 09SLH3 PRELIMINARY; PRT: 587 AA.

DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
DE PUTATIVE RGA1, GIBBERELLIN RESPONSE MODULATION PROTEIN.  
GN AT2G01570.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
NX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RX MEDLINE=20083487; PubMed=10617197;  
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,  
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,  
RA Cronin L.A., Shen M., Vanden S.E., Umayam L., Tallon L.J., Gill J.E.,  
RA Adams M.D., Carreira A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
RA Salzberg S.L., Fraser C.M., Venter J.C.;  
RT "Sequence and analysis of chromosome II of Arabidopsis thaliana.";  
RL Nature 402:761-768(1999).  
DR EMBL: AC005560; AAC67333.1;  
SQ SEQUENCE 587 AA; 64035 MW; FC92E7F9408072AA CRC64;

Query Match  
Best Local Similarity 3.0%; Score 19; DB 10; Length 587;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 CPYLKFAHFTANQALILEAF 332  
Db 299 CPYLKFAHFTANQALILEAF 317

RESULT 15  
ID 043476 PRELIMINARY; PRT: 275 AA.

DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE MAMMALIAN ACYL COA OXIDASE HOMOLOGOUS (FRAGMENT).  
GN CDR29.  
OS Hordeum vulgare (Barley).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
OC Triticeae; Hordeum.  
NX NCBI\_TaxID=4513;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. ARDA; TISSUE=GREEN LEAF;  
RA Grossi M., Galli M., Stanca A.M., Cattivelli L.;  
RL Plant Sci. 105:71-80(1995).  
DR EMBL: X84055; CAA58874.1;  
DR Mendel: 12863; Horvu:2063;12863.  
DR InterPro: IPR002655; ACOX.  
DR Pfam: PF01756; ACOX; 1.  
FT NON\_TER 1  
SQ SEQUENCE 275 AA; 30689 MW; 332811FD05827472 CRC64;

Query Match  
Best Local Similarity 2.6%; Score 16; DB 10; Length 275;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 482 FLDRTFESLHYSTMF 497  
Db 8 FLDRTFESLHYSTMF 23

Search completed: January 22, 2002, 16:36:44  
Job time: 344 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2002, 16:26:20 ; Search time 22.77 Seconds  
(without alignments)  
615.703 Million cell updates/sec

Title: US-09-485-529-7

Sequence: 1 MKREYDAGSGGGGGGMS.....TLGHTPLATSAWRLAGP 623

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 212252 seqs, 22503292 residues

Word size : 0

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents:AA:\*  
1: /cgn2.6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2.6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2.6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2.6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2.6/ptodata/2/1aa/PCITUS.COMB.pep:\*  
6: /cgn2.6/ptodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	3.0	259	4	US-09-117-853-6
2	19	3.0	282	4	US-09-117-853-8
3	19	3.0	532	4	US-09-117-853-2
4	14	2.2	221	4	US-09-117-853-4
5	12	1.9	475	2	US-08-861-464-14
6	12	1.9	475	2	US-08-396-001-14
7	12	1.9	475	4	US-09-323-433A-14
8	10	1.6	166	4	US-09-117-853-10
9	10	1.6	193	2	US-08-679-765-3
10	10	1.6	193	2	US-09-196-525-3
11	10	1.6	193	4	US-09-318-317-3
12	10	1.6	568	1	US-08-320-559-30
13	10	1.6	568	1	US-08-545-860D-30
14	10	1.6	568	5	PCR-US94-04496-30
15	10	1.6	590	2	US-08-785-310A-5
16	10	1.6	594	2	US-08-785-310A-6
17	9	1.4	11	1	US-08-785-310A-43
18	9	1.4	11	1	US-08-040-546-43
19	9	1.4	11	6	US-08-466-344-43
20	9	1.4	11	6	5206152-12
21	9	1.4	130	2	US-08-630-822A-90
22	9	1.4	130	2	US-09-005-068-90
23	9	1.4	148	1	US-08-207-904-15
24	9	1.4	180	6	5273901-7
25	9	1.4	180	6	5482709-6
26	9	1.4	199	3	US-08-506-553C-8
27	9	1.4	216	3	US-08-506-553C-23

28	9	1.4	218	4	US-09-068-655-7	Sequence 7, Appl
29	9	1.4	246	4	US-09-185-160-7	Sequence 7, Appl
30	9	1.4	262	1	US-08-106-981-6	Sequence 6, Appl
31	9	1.4	276	3	US-08-506-553C-26	Sequence 26, Appl
32	9	1.4	285	3	US-08-482-085B-20	Sequence 20, Appl
33	9	1.4	285	4	US-09-318-661-4	Sequence 4, Appl
34	9	1.4	298	4	US-09-318-661-2	Sequence 2, Appl
35	9	1.4	300	3	US-08-765-856-2	Sequence 2, Appl
36	9	1.4	300	4	US-08-935-009A-2	Sequence 2, Appl
37	9	1.4	302	3	US-08-765-856-4	Sequence 4, Appl
38	9	1.4	302	4	US-08-935-009A-4	Sequence 4, Appl
39	9	1.4	310	3	US-08-651-136C-22	Sequence 22, Appl
40	9	1.4	322	1	US-08-014-943A-2	Sequence 2, Appl
41	9	1.4	322	1	US-08-486-421-3	Sequence 3, Appl
42	9	1.4	322	1	US-08-470-911-3	Sequence 3, Appl
43	9	1.4	322	2	US-08-486-809-3	Sequence 3, Appl
44	9	1.4	345	2	US-08-282-197C-50	Sequence 50, Appl
45	9	1.4	360	3	US-08-899-437-7	Sequence 7, Appl

#### ALIGNMENTS

```
RESULT 1
US-09-117-853-6
; Sequence 6, Application US/09117853
; Patent No. 6307126
; GENERAL INFORMATION:
; APPLICANT: Harberd, Nicholas P
; APPLICANT: Peng, Jintong
; APPLICANT: Carol, Pierre
; APPLICANT: Richards, Donald E
; TITLE OF INVENTION: Nucleic acid encoding GAI gene of Arabidopsis thaliana
; FILE REFERENCE: 620-45
; CURRENT APPLICATION NUMBER: US/09/117,853
; CURRENT FILING DATE: 1998-08-12
; EARLIER APPLICATION NUMBER: PCT/GB97/00390
; EARLIER FILING DATE: 1997-02-12
; EARLIER APPLICATION NUMBER: GB 9602796.6
; EARLIER FILING DATE: 1996-02-12
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 6
; LENGTH: 259
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-117-853-6

Query Match 3.0%; Score 19; DB 4; Length 259;
Best Local Similarity 100.0%; Pred. No. 4.4e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 314 CPTLKFHFTANQATLEAF 332
DB 229 CPTLKFHFTANQATLEAF 247

RESULT 2
US-09-117-853-8
; Sequence 8, Application US/09117853
; Patent No. 6307126
; GENERAL INFORMATION:
; APPLICANT: Harberd, Nicholas P
; APPLICANT: Peng, Jintong
; APPLICANT: Carol, Pierre
; APPLICANT: Richards, Donald E
; TITLE OF INVENTION: Nucleic acid encoding GAI gene of Arabidopsis thaliana
; FILE REFERENCE: 620-45
; CURRENT APPLICATION NUMBER: US/09/117,853
; CURRENT FILING DATE: 1998-08-12
; EARLIER APPLICATION NUMBER: PCT/GB97/00390
; EARLIER FILING DATE: 1997-02-12
```

;; EARLIER APPLICATION NUMBER: GB 9602796.6  
;; EARLIER FILING DATE: 1996-02-12  
;; NUMBER OF SEQ ID NOS: 12  
;; SOFTWARE: Patentln Ver. 2.0  
;; SEQ ID NO 8  
;; LENGTH: 282  
;; TYPE: PRT  
;; ORGANISM: Arabidopsis thaliana  
US-09-117-853-8

Query Match 3.0%; Score 19; DB 4; Length 282;  
Best Local Similarity 100.0%; Pred. No. 4.8e-09;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 CPLYKFAHFTANQALILEAF 332  
|||||  
DB 229 CPLYKFAHFTANQALILEAF 247

RESULT 3  
US-09-117-853-2  
; Sequence 2, Application US/09117853  
; Patent No. 6307126  
; GENERAL INFORMATION:  
; APPLICANT: Harberd, Nicholas P  
; APPLICANT: Peng, Jinrong  
; APPLICANT: Carol, Pierre  
; APPLICANT: Richards, Donald E  
; TITLE OF INVENTION: Nucleic acid encoding GAI gene of Arabidopsis thaliana  
; FILE REFERENCE: 620-45  
; CURRENT APPLICATION NUMBER: US/09/117,853  
; CURRENT FILING DATE: 1998-08-12  
; EARLIER APPLICATION NUMBER: PCT/GB97/00390  
; EARLIER FILING DATE: 1997-02-12  
; EARLIER APPLICATION NUMBER: GB 9602796.6  
; EARLIER FILING DATE: 1996-02-12  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 532  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
US-09-117-853-2

Query Match 3.0%; Score 19; DB 4; Length 532;  
Best Local Similarity 100.0%; Pred. No. 9e-09;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 CPLYKFAHFTANQALILEAF 332  
|||||  
DB 246 CPLYKFAHFTANQALILEAF 264

RESULT 4  
US-09-117-853-4  
; Sequence 4, Application US/09117853  
; Patent No. 6307126  
; GENERAL INFORMATION:  
; APPLICANT: Harberd, Nicholas P  
; APPLICANT: Peng, Jinrong  
; APPLICANT: Carol, Pierre  
; APPLICANT: Richards, Donald E  
; TITLE OF INVENTION: Nucleic acid encoding GAI gene of Arabidopsis thaliana  
; FILE REFERENCE: 620-45  
; CURRENT APPLICATION NUMBER: US/09/117,853  
; CURRENT FILING DATE: 1998-08-12  
; EARLIER APPLICATION NUMBER: PCT/GB97/00390  
; EARLIER FILING DATE: 1997-02-12  
; EARLIER APPLICATION NUMBER: GB 9602796.6  
; EARLIER FILING DATE: 1996-02-12  
; NUMBER OF SEQ ID NOS: 12

;; SOFTWARE: Patentln Ver. 2.0  
;; SEQ ID NO 4  
;; LENGTH: 221  
;; TYPE: PRT  
;; ORGANISM: Arabidopsis thaliana  
US-09-117-853-4

Query Match 2.2%; Score 14; DB 4; Length 221;  
Best Local Similarity 100.0%; Pred. No. 0.00011;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 233 RLVHALLACAEAVQ 246  
|||||  
DB 150 RLVHALLACAEAVQ 163

RESULT 5  
US-08-861-464-14  
; Sequence 14, Application US/08861464  
; Patent No. 5874210  
; GENERAL INFORMATION:  
; APPLICANT: Guarente, Leonard P.  
; APPLICANT: Austriaco Jr., Nicanor  
; APPLICANT: Kennedy, Brian  
; TITLE OF INVENTION: Genes Determining Cellular Senescence  
; TITLE OF INVENTION: In Yeast  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Millia Drive  
; CITY: Lexington  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentln Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/861,464  
; FILING DATE: 22-MAY-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/396,001  
; FILING DATE: 28-FEB-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/09351  
; FILING DATE: 15-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/107,408  
; FILING DATE: 16-AUG-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Granahan, Patricia  
; REGISTRATION NUMBER: 32,227  
; REFERENCE/DOCKET NUMBER: MIT-6408A22  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 781-861-6240  
; TELEFAX: 781-861-9540  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 475 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-861-464-14

Query Match 1.9%; Score 12; DB 2; Length 475;  
Best Local Similarity 100.0%; Pred. No. 0.014;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 183 SSTSSSSSSSS 194  
 |||||||  
 DB 182 SSTSSSSSSSS 193

## RESULT 6

US-08-396-001-14  
 ; Sequence 14, Application US/08396001  
 ; Patent No. 5919618  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Guarente, Leonard P.  
 ; APPLICANT: Austriaco Jr., Nicanor  
 ; APPLICANT: Claus, James  
 ; APPLICANT: Cole, Francesca  
 ; APPLICANT: Kennedy, Brian  
 ; TITLE OF INVENTION: Genes Determining Cellular Senescence in  
 ; NUMBER OF SEQUENCES: 16  
 ; CORRESPONDENCE ADDRESSES:  
 ; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
 ; STREET: Two Militia Drive  
 ; CITY: Lexington  
 ; STATE: MA  
 ; COUNTRY: USA  
 ; ZIP: 02173  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/396,001  
 ; FILING DATE: 28-FEB-1995  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Granahan, Patricia  
 ; REGISTRATION NUMBER: 32,227  
 ; REFERENCE/DOCKET NUMBER: MIT-6408A2  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 617-861-6240  
 ; TELEFAX: 617-861-9940  
 ; INFORMATION FOR SEQ ID NO: 14:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 475 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-396-001-14

Query Match 1.9%; Score 12; DB 2; Length 475;  
 Best Local Similarity 100.0%; Pred. No. 0.014;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 183 SSTSSSSSSSS 194  
 |||||||  
 DB 182 SSTSSSSSSSS 193

RESULT 7  
 US-09-323-433A-14  
 ; Sequence 14, Application US/09323433A  
 ; Patent No. 6218512  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Guarente, Leonard P.  
 ; APPLICANT: Austriaco Jr., Nicanor  
 ; APPLICANT: Claus, James J.  
 ; APPLICANT: Cole, Francesca  
 ; APPLICANT: Kennedy, Brian  
 ; TITLE OF INVENTION: Genes DETERMINING CELLULAR SENESCENCE IN  
 ; NUMBER OF SEQUENCES: 5  
 ; FILE REFERENCE: 0050.1491-003  
 ; CURRENT APPLICATION NUMBER: US/09/323,433A

; CURRENT FILING DATE: 1999-06-01  
 ; PRIOR APPLICATION NUMBER: US 08/396,001  
 ; PRIOR FILING DATE: 1995-02-28  
 ; PRIOR APPLICATION NUMBER: PCT/US94/09351  
 ; PRIOR FILING DATE: 1994-08-15  
 ; PRIOR APPLICATION NUMBER: US 08/107,408  
 ; PRIOR FILING DATE: 1993-08-16  
 ; NUMBER OF SEQ ID NOS: 48  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 14  
 ; LENGTH: 475  
 ; TYPE: PRT  
 ; ORGANISM: Saccharomyces cerevisiae  
 ; US-09-323-433A-14

Query Match 1.9%; Score 12; DB 4; Length 475;  
 Best Local Similarity 100.0%; Pred. No. 0.014;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 183 SSTSSSSSSSS 194  
 |||||||  
 DB 182 SSTSSSSSSSS 193

RESULT 8  
 US-09-117-853-10  
 ; Sequence 10, Application US/09117853  
 ; Patent No. 6307126  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Harberd, Nicholas P  
 ; APPLICANT: Peng, Jintong  
 ; APPLICANT: Carol, Pierre  
 ; APPLICANT: Richards, Donald E  
 ; TITLE OF INVENTION: Nucleic acid encoding GAI gene of Arabidopsis thaliana  
 ; FILE REFERENCE: 620-45  
 ; CURRENT APPLICATION NUMBER: US/09/117,853  
 ; CURRENT FILING DATE: 1998-08-12  
 ; EARLIER APPLICATION NUMBER: PCT/GB97/00390  
 ; EARLIER FILING DATE: 1997-02-12  
 ; EARLIER APPLICATION NUMBER: GB 9602796.6  
 ; EARLIER FILING DATE: 1996-02-12  
 ; NUMBER OF SEQ ID NOS: 12  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 10  
 ; LENGTH: 166  
 ; TYPE: PRT  
 ; ORGANISM: Arabidopsis thaliana  
 ; US-09-117-853-10

Query Match 1.6%; Score 10; DB 4; Length 166;  
 Best Local Similarity 100.0%; Pred. No. 0.3;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 55 DVAKLEOLE 64  
 |||||||  
 DB 27 DVAKLEOLE 36

RESULT 9  
 US-08-679-765-3  
 ; Sequence 3, Application US/08679765  
 ; Patent No. 5840866  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Au-Young, Janice  
 ; APPLICANT: Goll, Surya K.  
 ; APPLICANT: Hillman, Jennifer L.  
 ; TITLE OF INVENTION: A NOVEL HUMAN UBQUITIN-CONJUGATING ENZYME  
 ; NUMBER OF SEQUENCES: 5  
 ; CORRESPONDENCE ADDRESSES:  
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
 ; STREET: 3174 Porter Drive

CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/679,765  
FILING DATE: Filed Herewith  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0093 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 193 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 1064914  
US-08-679-765-3

Query Match 1.6%; Score 10; DB 2; Length 193;  
Best Local Similarity 100.0%; Pred. No. 0.35;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 STSSSSSSSS 193  
|||||  
DB 9 STSSSSSSSS 18

RESULT 10  
US-09-196-525-3  
Sequence 3, Application US/09196525  
Patent No. 5989883  
GENERAL INFORMATION:  
APPLICANT: Au-Young, Janice  
APPLICANT: Goli, Surya K.  
APPLICANT: Hillman, Jennifer L.  
TITLE OF INVENTION: A NOVEL HUMAN UBIQUITIN-CONJUGATING  
TITLE OF INVENTION: ENZYME  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/196,525  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/679,765  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0093 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 193 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 1064914  
US-09-196-525-3

Query Match 1.6%; Score 10; DB 2; Length 193;  
Best Local Similarity 100.0%; Pred. No. 0.35;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 STSSSSSSSS 193  
|||||  
DB 9 STSSSSSSSS 18

RESULT 11  
US-09-318-317-3  
Sequence 3, Application US/09318317  
Patent No. 6172199  
GENERAL INFORMATION:  
APPLICANT: Au-Young, Janice  
APPLICANT: Goli, Surya K.  
APPLICANT: Hillman, Jennifer L.  
TITLE OF INVENTION: A NOVEL HUMAN UBIQUITIN-CONJUGATING  
TITLE OF INVENTION: ENZYME  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/318,317  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/196,525  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0093 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 193 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 1064914  
US-09-318-317-3

Query Match 1.6%: Score 10; DB 4; Length 193;

Best Local Similarity 100.0%; Pred. No. 0.35; Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 184 TSSSSSSSS 193

Db 9 STSSSSSSSS 18

RESULT 12

US-08-320-559-30

; Sequence 30, Application US/08320559

; Patent No. 5633135

; GENERAL INFORMATION:

; APPLICANT: Croce, Carlo

; APPLICANT: Canaan, Eli

; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods for

; TITLE OF INVENTION: Detection and Treatment of Acute Leukemias

; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the

; TITLE OF INVENTION: All-1 Region

; NUMBER OF SEQUENCES: 44

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5633135rls

; STREET: One Liberty Place - 46th Floor

; CITY: Philadelphia

; STATE: PA

; COUNTRY: USA

; ZIP: 19103

; COMPUTER READABLE FORM:

; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE

; OPERATING SYSTEM: PC-DOS

; SOFTWARE: WORDPERFECT 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/320,559

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/062,443

; FILING DATE: 14 MAY 1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/07/971,094

; FILING DATE: 30-OCT-92

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/07/888,830

; FILING DATE: 27-MAY-92

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/07/805,093

; FILING DATE: 11-DEC-91

; ATTORNEY/AGENT INFORMATION:

; NAME: Deluca, Mark

; REGISTRATION NUMBER: 33,229

; REFERENCE/DOCKET NUMBER: TJU-0855

; TELEPHONE: (215) 568-3100

; TELEFAX: (215) 568-3439

; INFORMATION FOR SEQ ID NO: 30:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 568 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-320-559-30

Query Match 1.6%: Score 10; DB 1; Length 568;

Best Local Similarity 100.0%; Pred. No. 1; Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 185 TSSSSSSSS 194

|||||||

Db 148 TSSSSSSSS 157

RESULT 13

US-08-545-860D-30

; Sequence 30, Application US/08545860D

; Patent No. 6040140

; GENERAL INFORMATION:

; APPLICANT: Croce, Carlo

; APPLICANT: Canaan, Eli

; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods

; TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias

; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 Region

; NUMBER OF SEQUENCES: 94

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &

; ADDRESS: No. 6040140rls

; STREET: One Liberty Place, 46th floor

; CITY: Philadelphia

; STATE: Pennsylvania

; COUNTRY: USA

; ZIP: 19103

; COMPUTER READABLE FORM:

; MEDIUM TYPE: FLOPPY disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/545,860D

; FILING DATE: 07-MAR-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US94/04496

; FILING DATE: 22-APR-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US92/10930

; FILING DATE: 09-DEC-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/320,559

; FILING DATE: 11-OCT-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/062,443

; FILING DATE: 14-MAY-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/971,094

; FILING DATE: 30-OCT-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/888,839

; FILING DATE: 27-MAY-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/805,093

; FILING DATE: 11-DEC-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Deluca Esq., Mark

; REGISTRATION NUMBER: 33,229

; REFERENCE/DOCKET NUMBER: TJU-1262

; TELEPHONE: (215) 568-3100

; TELEFAX: (215) 568-3439

; INFORMATION FOR SEQ ID NO: 30:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 568 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-545-860D-30

Query Match 1.6%: Score 10; DB 3; Length 568;

Best Local Similarity 100.0%; Pred. No. 1;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 TSSSSSSSS 194  
Db 148 TSSSSSSSS 157

## RESULT 14

PCT-US94-04496-30  
Sequence 30, Application PC/TUS9404496  
GENERAL INFORMATION:  
APPLICANT: Croce, Carlo  
APPLICANT: Canaan, Eli  
TITLE OF INVENTION: Diagnostics, Therapeutics and Methods  
TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias  
TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the ALL-1  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock, Washburn, Kurtz, Macklewitz &  
ADDRESSEE: Norris  
STREET: One Liberty Place, 46th floor  
CITY: Philadelphia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/04496  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Deluca Esq., Mark  
REGISTRATION NUMBER: 33,229  
REFERENCE/DOCKET NUMBER: TJU-1242  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 568 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US94-04496-30

Query Match 1.6%; Score 10; DB 5; Length 568;  
Best Local Similarity 100.0%; Pred. No. 1;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 TSSSSSSSS 194  
Db 148 TSSSSSSSS 157

## RESULT 15

US-08-785-310A-5  
Sequence 5, Application US/08785310A  
Patent No. 5840532  
GENERAL INFORMATION:  
APPLICANT: McKnight, Steven L.  
APPLICANT: Russell, David W.  
TITLE OF INVENTION: Neuronal PAS Domain Protein  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
STREET: 268 BUSH STREET, SUITE 3200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA

COUNTRY: USA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/785,310A  
FILING DATE: 21-JAN-1997  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: OSMAN, RICHARD A  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: UTSD:1226  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 343-4341  
TELEFAX: (415) 343-4342  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 590 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-785-310A-5

Query Match 1.6%; Score 10; DB 2; Length 590;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 SSSSSSSSL 195  
Db 214 SSSSSSSSL 223

Search completed: January 22, 2002, 16:30:53  
Job time: 273 sec



---





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 23, 2002, 02:37:31 ; Search time 3020.32 Seconds  
(Without alignments)  
11606.885 Million cell updates/sec

Title: US-09-485-529-14

Sequence: 1 atagagagcgaggtagctc.....tggatggacgacgaactccg 2125

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba: \*  
2: gb\_htg: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sts: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vi: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_om: \*  
20: em\_or: \*  
21: em\_ov: \*  
22: em\_pat: \*  
23: em\_ph: \*  
24: em\_pl: \*  
25: em\_ro: \*  
26: em\_sts: \*  
27: em\_sy: \*  
28: em\_un: \*  
29: em\_vi: \*  
30: em\_htgo\_hum: \*  
31: em\_htgo\_inv: \*  
32: em\_htgo\_rod: \*  
33: em\_htg\_hum: \*  
34: em\_htg\_inv: \*  
35: em\_htg\_rod: \*  
36: em\_htg\_other: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2125	100.0	2125	6	AX005805	AX005805 Sequence
2	1821	85.7	1872	6	TAE242531	AJ242531 Triticum
3	1453	68.4	1768	6	AX005804	AX005804 Sequence
4	659	31.0	2709	6	AX005794	AX005794 Sequence
5	652	30.7	1746	6	AX005848	AX005848 Sequence
6	280	13.2	377	6	AX005862	AX005862 Sequence
7	251	11.8	453	6	AX005857	AX005857 Sequence
8	217	10.2	416	6	AX005809	AX005809 Sequence
9	210	9.9	425	6	AX005864	AX005864 Sequence
10	159	7.5	332	6	AX005849	AX005849 Sequence
11	157	7.4	324	6	AX005867	AX005867 Sequence
12	146	6.9	436	6	AX005863	AX005863 Sequence
13	122	5.7	511	6	AX005854	AX005854 Sequence
14	107	5.0	357	6	AX005853	AX005853 Sequence
15	94	4.4	230	6	AX005861	AX005861 Sequence
16	89	4.2	369	6	AX005852	AX005852 Sequence
17	81	3.8	211	6	AX005851	AX005851 Sequence
18	76	3.6	309	6	AX005855	AX005855 Sequence
19	65	3.1	1890	8	ZMA242530	AJ242530 Zea mays
20	65	3.1	2255	6	AX005806	AX005806 Sequence
21	57	2.7	285	6	AX005865	AX005865 Sequence
22	53	2.5	2500	8	AB030956	AB030956 Oryza sat
23	53	2.5	122497	8	AC087797	AC087797 Oryza sat
24	52	2.4	371	6	AX005808	AX005808 Sequence
25	52	2.4	377	6	AX005850	AX005850 Sequence
26	51	2.4	51	6	AX005896	AX005896 Sequence
27	48	2.3	259	6	AX005866	AX005866 Sequence
28	48	2.3	399	6	AX005856	AX005856 Sequence
29	45	2.1	800	8	AF377621	AF377621 Zea mays
30	45	2.1	800	8	AF377622	AF377622 Zea mays
31	45	2.1	800	8	AF377623	AF377623 Zea mays
32	45	2.1	800	8	AF377624	AF377624 Zea mays
33	45	2.1	800	8	AF377625	AF377625 Zea mays
34	45	2.1	800	8	AF377626	AF377626 Zea mays
35	45	2.1	800	8	AF377627	AF377627 Zea mays
36	45	2.1	800	8	AF377628	AF377628 Zea mays
37	45	2.1	800	8	AF377629	AF377629 Zea mays
38	45	2.1	800	8	AF377630	AF377630 Zea mays
39	45	2.1	800	8	AF377631	AF377631 Zea mays
40	45	2.1	800	8	AF377632	AF377632 Zea mays
41	45	2.1	800	8	AF377633	AF377633 Zea mays
42	45	2.1	800	8	AF377634	AF377634 Zea mays
43	45	2.1	800	8	AF377635	AF377635 Zea mays
44	45	2.1	800	8	AF377636	AF377636 Zea mays
45	45	2.1	800	8	AF377637	AF377637 Zea mays

#### ALIGNMENTS

RESULT 1  
LOCUS AX005805  
DEFINITION Sequence 14 from Patent WO9909174.  
ACCESSION AX005805  
VERSION AX005805.1 GI:9928802  
KEYWORDS  
SOURCE  
ORGANISM  
Triticum aestivum  
bread wheat.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Triticum.  
REFERENCE  
AUTHORS Harberd N.P. and Peng J.  
TITLE Genetic control of plant growth and development  
JOURNAL Patent: WO 9909174-A 14 25-FEB-1999;  
HARBERD NICHOLAS PAUL (GB); PENG JINRONG (GB)  
FEATURES  
1..2125  
/organism="Triticum aestivum"



	Db	2041	GAAATTCAGCGACGCCGCTACATGCATGGGTGGCCTACGGTGGTTCCGTCGCTTGCGCTGA	2100
Oy	2101	agaagtgatgacgaactccg	2125	
Db	2101	AGAGGTGATGACGACGACTCG	2125	
RESULT	2	TAE242531		
LOCUS		TAE242531	1872 bp DNA	PLN 28-JUL-1999
DEFINITION		Triticum aestivum rht-D1a gene for gibberellin response modulator.		
ACCESSION		AJ242531		
VERSION		AJ242531.1	GI:5640156	
KEYWORDS		gibberellin response modulator; rht-D1a gene.		
SOURCE		bread wheat.		
ORGANISM		Triticum aestivum		
REFERENCE		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
AUTHORS		Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;		
		Poideae; Triticeae; Triticum.		
TITLE		1 (bases 1 to 1872)		
JOURNAL		Peng,J., Richards,D.E., Hartley,N.M., Murphy,G.P., Devos,K.M.,		
MEDLINE		Flintham,J.E., Beales,J., Fish,L.J., Worland,A.J., Pelica,F.,		
REFERENCE		Sudharak,D., Christou,P., Snape,J.W., Gale,M.D., and Harberd,N.P.		
AUTHORS		'Green revolution' genes encode mutant gibberellin response		
TITLE		modulators		
NATURE		400 (6741), 256-261 (1999)		
99347734				
2 (bases 1 to 1872)				
Harberd,N.P., Peng,J. and Richards,D.E.				
Green revolution genes encode mutant gibberellin response				
modulators				
Unpublished				
3 (bases 1 to 1872)				
Richards,D.E.				
Direct Submission				
Submitted (25-MAY-1999) Richards D.E., Molecular Genetics, John				
Innes Centre, Colney lane, NR4 7UJ, UNITED KINGDOM				
Location/Qualifiers				
1..1872		/organism="Triticum aestivum"		
/db_xref="taxon:4565"				
1..1872				
/gene="rht-D1a"				
1..1872		/gene="rht-D1a"		
/note="GAI ortholog"				
/codon_start=1				
/product="gibberellin response modulator"				
/protein_id="CA5I1555.1"				
/db_xref="GI:5640157"				
/translation="MKREYODAGSGGGGGMGSSSEDKMVSAAAGEEVEDLLAAL				
GKRAISLADMDVAOKLEOLEAMGMGVGAAGAADDSPATHLDVTHYNPTIDPSMW				
ESMSSELNPAPPPRLPAPOLNASTSRVTGSQGYFDLPVSDSSSTIALRPIIPSPA				
ATAPADVDASDVRPKPMRKRTGSSSTSSSSSLGGARBSVTEAAPPAVAANAATTPE				
ALPVAVDTDGRIQLVALHACAEAVQEQENLSAELVKOITPLLASQGAMRVAAV				
YCFGNARVFRRFRPOPSSDLDAFADLILAHRYESCPLYKFAPHANOAIIIEAFNG				
CRRHVAVDFIKFGKMOMPALLOALPRPGSPSRLGVGPDPDETDALQQVMKTLAG				
OFATTVIRVDFOREGIVNAATLADIEDPFMLQPEGEPNEPEVIADVNSVEFHRLIAO				
GALEXKVGTVAVRPRIATVVYEQGANHNNSCRFLDRFTESIHYTSMTDSLEGSSGGG				
PSKVSSCAAAPAAGTDOWSERYTLRGOTCNVACBEATEKHETLGQRNRNLGN				
GFETVLHGSNAYKKASLTLLALEFAGGDGYKEEREGCLTLGWHTPLLATISAWRLAGP				
BASE COUNT	265	a	700	c 641 g 266 t
ORIGIN				
Query Match	85.7%	Score 1821:	DB 8:	Length 1872:
Best Local Similarity	99.9%:	Pred. No. 0;		
Matches 1871: Conservative	0:	Mismatches	1;	Indels 0; Gaps 0;
Oy	28	atgaaggcggtattaccagcgcgcggagggcgcgcgcgctgtgcgcgatgacctcg	87	
Db	1	ATGAAGCGGGAATTACCAGACGCCCGAGGACCGCGCGCGGTGGCGCATGGAGCTCG	60	

[illegible]

[illegible][illegible]



Db 950 CAGGGGCGGCGATGCGGAGAGTCCGCGCTACTTCGGGAGAGCCCTCGCCCGCGCTC 1009  
QY 886 ttccgcttcgcccgcgagcgagcagctccctccctcagcgcgaccttcgacgacctc 945  
Db 1010 TTCGGTTCCGCGCCGAGCGGACAGCTCCCTCTCTGAGCGCCGCTTCGCGCACTCTC 1069  
QY 946 cagcgacgttcacagatcgtgccccctacacacacacacacacacacacacacacacac 1005  
Db 1070 CACGCGCACTTACAGAGTCCGCGCCCTACCTCAAGTTGCGCAGCTTCACCGCAACAG 1129  
QY 1006 gcatctctgaaagcgttcgcccgcgagcgagcgagcgagcgagcgagcgagcgagcgag 1065  
Db 1130 GCGATCTGGAAGCTGCTCCGCGGCTCCGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 1189  
QY 1066 cagggagatcagtgcccgcgacgttcacacacacacacacacacacacacacacac 1125  
Db 1190 CAGGGGATCAGATGCGCGGCACTTCTCCAGGCGCTCGCGCTCGCGCTCGCGCTCGCG 1249  
QY 1126 tgcgttcgacctcagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1185  
Db 1250 TCGTTCCGCTCAGCGGCTCGCGGCTCGCGGCTCGCGGCTCGCGGCTCGCGGCTCGCG 1309  
QY 1186 gtaggctgaaagcgttcgcccgcgagcgagcgagcgagcgagcgagcgagcgagcgag 1245  
Db 1310 GAGGCTGGAAGCTGCGGCTCGCGGCTCGCGGCTCGCGGCTCGCGGCTCGCGGCTCGCG 1369  
QY 1246 ctgcgttcgacctcagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1304  
Db 1370 CTCGTCGCGGCTCGCGGCTCGCGGCTCGCGGCTCGCGGCTCGCGGCTCGCGGCTCGCG 1428

## RESULT 5

AX005848 1746 bp DNA PAT 24-AUG-2000  
LOCUS AX005848  
DEFINITION Sequence 57 from Patent WO9909174.  
ACCESSION AX005848  
VERSION AX005848.1 GI:9928843  
KEYWORDS  
SOURCE  
ORGANISM  
Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Triticum.  
REFERENCE  
1 (bases 1 to 1746)  
AUTHORS Harberd,N.P. and Peng,J.  
TITLE Genetic control of plant growth and development  
JOURNAL Patent: WO 9909174-A 57 25-FEB-1999;  
HARBERD NICHOLAS PAUL (GB); PENG JINRONG (GB)  
FEATURES  
Location/Qualifiers  
source  
1..1746  
/organism="Triticum aestivum"  
/db\_xref="taxon:4565"

misc\_feature  
/note="n is any nucleotide"  
680  
misc\_feature  
/note="n is any nucleotide"  
786  
misc\_feature  
/note="n is any nucleotide"  
972  
misc\_feature  
/note="n is any nucleotide"  
1038  
misc\_feature  
/note="n is any nucleotide"  
1130  
misc\_feature  
/note="n is any nucleotide"  
1170  
misc\_feature  
/note="n is any nucleotide"  
1212  
BASE COUNT 375 a 590 c 495 g 278 t 8 others  
ORIGIN

Query Match 30.7%; Score 652; DB 6; Length 1746;

Best Local Similarity 100.0%; Pred. No. 9.9e-291;  
Matches 652; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 653 cgttcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 712  
Db 7 CCGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 66  
QY 713 aggcgagatcagtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 772  
Db 67 AGGCGGAGTTGGCGGAGTGGCGGAGTGGCGGAGTGGCGGAGTGGCGGAGTGGCGGAGTGG 126  
QY 773 acccttcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 832  
Db 127 ACCTTCGCGCGCGGAGGCGGAGTGGCGGAGTGGCGGAGTGGCGGAGTGGCGGAGTGGCG 186  
QY 833 ggcgagtgcgagagtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 892  
Db 187 GCGGAGTGGCGGAGTGGCGGAGTGGCGGAGTGGCGGAGTGGCGGAGTGGCGGAGTGGCGG 246  
QY 893 tccgcgcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 952  
Db 247 TCGGCGCGGAGCGGAGTGGCGGAGTGGCGGAGTGGCGGAGTGGCGGAGTGGCGGAGTGGCG 306  
QY 953 accttcagagtcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1012  
Db 307 ACTTCTACAGTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 366  
QY 1013 tggagagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1072  
Db 367 TGGAGGCGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 426  
QY 1073 tgcagtgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1132  
Db 427 TCGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 486  
QY 1133 gcttcagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1192  
Db 487 GCTCTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 546  
QY 1193 ggaagctgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1252  
Db 547 GGAAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 606  
QY 1253 ccgcagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1304  
Db 607 CCGCACGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 658

## RESULT 6

AX005862 377 bp DNA PAT 24-AUG-2000  
LOCUS AX005862  
DEFINITION Sequence 71 from Patent WO9909174.  
ACCESSION AX005862  
VERSION AX005862.1 GI:9928857  
KEYWORDS  
SOURCE  
ORGANISM  
Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Triticum.  
REFERENCE  
1 (bases 1 to 377)  
AUTHORS Harberd,N.P. and Peng,J.  
TITLE Genetic control of plant growth and development  
JOURNAL Patent: WO 9909174-A 71 25-FEB-1999;  
HARBERD NICHOLAS PAUL (GB); PENG JINRONG (GB)  
FEATURES  
Location/Qualifiers  
source  
1..377  
/organism="Triticum aestivum"  
/db\_xref="taxon:4565"

misc\_feature  
/note="n is any nucleotide"  
2  
misc\_feature  
/note="n is any nucleotide"  
3  
misc\_feature  
/note="n is any nucleotide"







REFERENCE	1 (bases 1 to 324)			
AUTHORS	Harberd,N.P. and Peng,J.			
TITLE	Genetic control of plant growth and development			
JOURNAL	Patent: WO 9909174-A 76 25-FEB-1999;			
	NICHOLAS PAUL (GB); PENG JINRONG (GB)			
FEATURES	Location/Qualifiers			
source	1..324			
	/organism="Triticum aestivum"			
	/db_xref="taxon:4565"			
misc_feature	158			
	/note="n is any nucleotide"			
misc_feature	161			
	/note="n is any nucleotide"			
misc_feature	217			
	/note="n is any nucleotide"			
BASE COUNT	39 a 141 c 85 g 56 t 3 others			
ORIGIN				
Query Match	7.4%; Score 157; DB 6; Length 324;			
Best Local Similarity	100.0%; Freq. NO. 1.0e-61;			
Matches 157; Conservative	0; Mismatches 0; Indels 0; Gaps 0;			
QY	144	gctctgagcagcgtctcggtatacaagttgcgagcccccagacatgagcgagcagttgcgcaaga	203	
	145		204	
Db	157	gctctgagcagcgtctcggtatcaaaagttgcgagcccccagacatgagcgagcagttgcgcaaga	98	
QY	204	gcttgagcagcgtctcgatgatggtccatggtggatggtgcggtgtggcgccgcgcgcgcgcga	263	
	205		264	
Db	97	gcttgagcagcgtctcgatgagagtgccatggggatgagcgcgctggcgccgcgcgcgcgcga	38	
QY	264	cgacagcttcgcacacccacccatcgccacgaggaacgctg	300	
	265		301	
Db	37	CGACAGCTTCGCCACCCACCTCGCCACGAGACCGCTG	1	
RESULT 12				
LOCUS	AX005863 436 bp DNA PAT 24-AUG-2000			
DEFINITION	Sequence 72 from Patent WO9909174.			
ACCESSION	AX005863			
VERSION	AX005863.1 GI:9928858			
KEYWORDS				
SOURCE	bread wheat.			
ORGANISM	Triticum aestivum			
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;			
	Pooidae; Triticeae; Triticum.			
REFERENCE	1 (bases 1 to 436)			
AUTHORS	Harberd,N.P. and Peng,J.			
TITLE	Genetic control of plant growth and development			
JOURNAL	Patent: WO 9909174-A 72 25-FEB-1999;			
	NICHOLAS PAUL (GB); PENG JINRONG (GB)			
FEATURES	Location/Qualifiers			
source	1..436			
	/organism="Triticum aestivum"			
	/db_xref="taxon:4565"			
misc_feature	1			
	/note="n is any nucleotide"			
misc_feature	21			
	/note="n is any nucleotide"			
misc_feature	54			
	/note="n is any nucleotide"			
misc_feature	60			
	/note="n is any nucleotide"			
misc_feature	225			
	/note="n is any nucleotide"			
misc_feature	267			
	/note="n is any nucleotide"			
misc_feature	364			

	misc_feature	/note="n is any nucleotide"	
	385	/note="n is any nucleotide"	
	misc_feature	/note="n is any nucleotide"	
	394	/note="n is any nucleotide"	
	404	/note="n is any nucleotide"	
	misc_feature	/note="n is any nucleotide"	
	412	/note="n is any nucleotide"	
	misc_feature	/note="n is any nucleotide"	
	415	/note="n is any nucleotide"	
	misc_feature	/note="n is any nucleotide"	
	416	/note="n is any nucleotide"	
	misc_feature	/note="n is any nucleotide"	
	425	/note="n is any nucleotide"	
	misc_feature	/note="n is any nucleotide"	
	426	/note="n is any nucleotide"	
	misc_feature	/note="n is any nucleotide"	
	427	/note="n is any nucleotide"	
	misc_feature	/note="n is any nucleotide"	
	433	/note="n is any nucleotide"	
	misc_feature	/note="n is any nucleotide"	
	436	/note="n is any nucleotide"	
	BASE COUNT	58 a 153 c 141 g 66 t 18 others	
	ORIGIN		
	Query Match	6.9%; Score 146; DB 6; Length 436;	
	Best Local Similarity	100.0%; Pred. NO. 1,8e-56;	
	Matches 146; Conservative	0; Mismatches 0; Indels 0; Gaps 0	
Oy	454 gtgcactcttcacagcagcatcaagcgctgaggcgatccctccggcgcgagcg 513		
	gtccc		
Db	76 gtgcacctcttccagcacacatctacggcgctgccccctcccgcgcgacg 135		
Oy	514 gcggcgcgcgacactgcttcgccgaactctcgtggcgagatcccaaaggatcgacatggcgag 573		
	gcggcgcgcgacactgcttcgccgaactctcgtggcgagatcccaaaggatcgacatggcgag		
Db	136 gcggcgcgcgacactgcttcgccgaactctcgtggcgagatcccaaaggatcgacatggcgag 195		
Oy	574 agcagcaccttgctgcatcctctc 599		
	gtccc		
Db	196 AGCAGCACTGTCGTGCATCTCCTC 221		
	RESULT 13		
	AX005854		
LOCUS	AX005854 511 bp DNA PAT 24-AUG-2000		
DEFINITION	Sequence 63 from Patent WO9909174.		
ACCESSION	AX005854		
VERSION	AX005854.1 GI:9928849		
KEYWORDS	bread wheat.		
SOURCE	Triticum aestivum		
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;		
	Poideae; Triticaceae; Triticum.		
REFERENCE	1 (bases 1 to 511)		
AUTHORS	Harberd,N.P. and Peng,J.		
TITLE	Genetic control of plant growth and development		
JOURNAL	Patent: WO 9909174-A 63 25-FEB-1999;		
	HARBERD NICHOLAS PAUL (GB); PENG JINRONG (GB)		
FEATURES	location/Oligoflucere		
source	1..511		
	/organism="Triticum aestivum"		
	/db_xref="taxon:4365"		
	7		
misc_feature	/note="n is any nucleotide"		
	17		
misc_feature	/note="n is any nucleotide"		
	149		
misc_feature	/note="n is any nucleotide"		
	263		
misc_feature	/note="n is any nucleotide"		

misc_feature	361	/note="n is any nucleotide"	
misc_feature	362	/note="n is any nucleotide"	
misc_feature	413	/note="n is any nucleotide"	
misc_feature	438	/note="n is any nucleotide"	
misc_feature	479	/note="n is any nucleotide"	
misc_feature	484	/note="n is any nucleotide"	
misc_feature	491	/note="n is any nucleotide"	
misc_feature	496	/note="n is any nucleotide"	
misc_feature	510	/note="n is any nucleotide"	
BASE COUNT	125 a	137 c	117 g
ORIGIN			13 others

Query Match	5.7%	Score 122;	DB 6;	Length 511;
Best Local Similarity	100.0%	Pred. No. 2.3e-45;		
Matches 122; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

Dy 1883 gccctggccggycccgatcctgcgaagttttaaagcgttaagtacacatcgtgaagcatcgg 194

Dd 18 gcttgccggggccggtgatctcgcgattttgaacgctgtaatcacatcgtgagcatcgg 77

Qy	1943	agagcaaacacagagcccgcgcgcgcccgctctccgycgaagcagcagcagcact	2002
Db	78	AGGACACACACAGCCCCGGGGCCGGCCGGCTCTCGGGGGAAGCAGCAGCAGCAGCACT	137

QY	2003	tg	2004
Db	138	TG	139

RESULT 14			
AX005853	AX005853	357 bp	DNA
LOCUS	Sequence 62 from Patent WO9909174.		PAT
DEFINITION	AX005853		24-AUG-2000
ACCESSION	AX005853.1		
VERSION	GI:9928848		
KEYWORDS			
SOURCE	bread wheat.		

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES

1 (bases: 1 to 357)  
Harberd, N. P. and Peng, J.  
Genetic control of plant growth and development  
Patent: WO 9509174-A 62 25-FEB-1995;  
HARBERD NICHOLAS PAUL (GB); PENG JINKONG (GB)  
Location/Qualifiers

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poaceae; Triticeae; Triticum.

```
misc_feature
6 /organism="Trifolium aestivum"
/db_xref="taxon:4565"
/misc_feature
57 /note="n is any nucleotide"
/misc_feature
177 /note="n is any nucleotide"
/misc_feature
235 /note="n is any nucleotide"
/misc_feature
238 /note="n is any nucleotide"
/misc_feature
251 /note="n is any nucleotide"
/misc_feature
286 /note="n is any nucleotide"
```

misc_feature	/note="n is any nucleotide"	293			
misc_feature	/note="n is any nucleotide"	294			
misc_feature	/note="n is any nucleotide"	295			
misc_feature	/note="n is any nucleotide"	296			
misc_feature	/note="n is any nucleotide"	328			
misc_feature	/note="n is any nucleotide"	331			
misc_feature	/note="n is any nucleotide"	332			
misc_feature	/note="n is any nucleotide"	333			
misc_feature	/note="n is any nucleotide"				
BASE COUNT	57 a	125 c	100 g	60 t	15 others
ORIGIN					

Query Match	5.0%;	Score 107;	DB 6;	Length 357;
Best Local Similarity	100.0%;	Pred. No. 2.1e-38;		
Matches 107;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Dy 1501 tactactccaccatgttctcgattcccgcgagggcgcgcagctccgcgcgcgccatccga 1560  
|||||  
Dd 58 TACTACTCCACCATGTTCGATTCCTCGAGGGCGGCAGCTCCGGCGGCCCATCCGAA 117

QY	1561 gtcacatcggggtctgtctgtcctcgccgcgcgcgcagcacca	160
Db	118 GTCTCATCGGGGCTGTGCTGCTCTCTCCGCCGCCGCAGCAGACA	164

RESULT	15
AX005861	
LOCUS	AX005861 230 bp DNA
DEFINITION	Sequence 70 from Patent WO9909174.
ACCESSION	AX005861
VERSION	AX005861.1 GI:3928856
KEYWORDS	.
SOURCE	bread wheat.
ORGANISM	Triticum aestivum L.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES

Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:  
Spermatophyta: Magnoliophyta: Liliopsida: Poales; Poaceae;  
1 (bases 1 to 230)  
Poideae: Triticeae: Triticum.  
Harberd N P and Peng J  
Genetic control of plant growth and development  
Patent: WO 9009174-A 70 25-FEB-1999.  
HARBERD NICHOLAS PAUL, (GB), PENG JINKONG (GB)  
Location/Qualifiers  
51  
52

	BASE COUNT	ORIGIN
misc_feature	4	14
misc_feature	14	202
misc_feature	202	210
misc_feature	210	214
misc_feature	214	28 t
misc_feature	28 t	5 others

Query Match	4.4%	Score 94;	DB 6;	Length 230;
Best Local Similarity	100.0%	Pred. No. 2.4e-32;		
Matches 94; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

QY	205	ctggagcagctcagatggccatgggagatggcgcgctggcgccgcccgcgac	264
Db	54	CTGGAGCAGCTCGAGATGGCCATGGGGATGGCGCGCTGGCGCCGCGCGCCCGAC	113
QY	265	gacagcttcgcaaccacctcgccacgacacg	298
Db	114	GACAGCTTCGCGCACCCACCTCGCCACGACACCG	147

Search completed: January 23, 2002, 04:54:37  
job time: 8226 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 23, 2002, 02:47:47 ; Search time 197.61 Seconds  
(without alignments)  
9219.251 Million cell updates/sec

Title: US-09-485-529-14  
Perfect score: 2125  
Sequence: 1 atagagagcgagtagctc.....tgagtgacgacgactccg 2125

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

N.Geneseq\_1101:\*

1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT:\*  
2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT:\*  
3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT:\*  
4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT:\*  
5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT:\*  
6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT:\*  
7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT:\*  
8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT:\*  
9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT:\*  
10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT:\*  
11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT:\*  
12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT:\*  
13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT:\*  
14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT:\*  
15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT:\*  
16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT:\*  
17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT:\*  
18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT:\*  
19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT:\*  
20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT:\*  
21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT:\*  
22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2125	100.0	2125	20	AA36279	Wheat Rht clone 5a
2	1453	68.4	1768	20	AA36278	Wheat Rht clone C1
3	659	31.0	2709	20	AA36275	Composite DNA sequ
4	652	30.7	1746	20	AA36285	Consensus cDNA seq
5	280	13.2	377	20	AA36268	DNA sequence obtai
6	251	11.8	453	20	AA36263	DNA sequence obtai
7	217	10.2	416	20	AA36283	Partial sequence o
8	210	9.9	425	20	AA36270	DNA sequence obtai
9	159	7.5	332	20	AA36255	DNA sequence obtai
10	157	7.4	324	20	AA36273	DNA sequence obtai
11	146	6.9	436	20	AA36269	DNA sequence obtai

12	122	5.7	511	20	AA36260	DNA sequence obtai
13	107	5.0	357	20	AA36259	DNA sequence obtai
14	94	4.4	360	20	AA36267	DNA sequence obtai
15	89	4.2	369	20	AA36258	DNA sequence obtai
16	81	3.8	211	20	AA36257	DNA sequence obtai
17	76	3.6	309	20	AA36261	DNA sequence obtai
18	65	3.1	2255	20	AA36280	Maize 1a1 genomic
19	57	2.7	285	20	AA36271	DNA sequence obtai
20	52	2.4	371	20	AA36282	Partial sequence o
21	52	2.4	377	20	AA36256	DNA sequence obtai
22	51	2.4	51	20	AA36284	Oligonucleotide de
23	48	2.3	259	20	AA36272	DNA sequence obtai
24	48	2.3	399	20	AA36262	DNA sequence obtai
25	44	2.1	302	20	AA36281	Partial sequence o
26	40	1.9	200	20	AA36266	DNA sequence obtai
27	35	1.6	35	20	AA36203	Primer used for se
28	35	1.6	35	20	AA36204	Primer used for se
29	35	1.6	770	20	AA36277	Rice EST D39460 se
30	34	1.6	725	20	AA36276	Partial cDNA sequ
31	27	1.3	27	20	AA36244	Primer used for se
32	27	1.3	35	20	AA36202	Primer used for se
33	27	1.3	2151	18	AA95767	Maize ZCARECROW ZC
34	27	1.3	2151	21	AA65298	Maize ZCR gene par
35	27	1.3	3510	21	AA65299	Maize Scarecrow nu
36	25	1.2	25	20	AA36233	Primer used for se
37	25	1.2	25	20	AA36214	Primer used for se
38	25	1.2	25	20	AA36226	Primer used for se
39	25	1.2	25	20	AA36213	Primer used for se
40	24	1.1	24	20	AA36215	Primer used for se
41	23	1.1	23	20	AA36231	Primer used for se
42	23	1.1	23	20	AA36232	Primer used for se
43	23	1.1	23	20	AA36228	Primer used for se
44	23	1.1	23	20	AA36229	Primer used for se
45	23	1.1	23	20	AA36207	Primer used for se

#### ALIGNMENTS

RESULT 1	
ID	AA36279 standard; DNA: 2125 BP.
XX	XX
AC	AA36279;
XX	XX
DT	16-JUL-1999 (first entry)
XX	XX
DE	Wheat Rht clone 5a1 genomic sequence.
XX	XX
KW	Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;
KW	antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;
KW	paclobutrazol; ss.
XX	XX
OS	Triticum aestivum.
XX	XX
PN	MO9909174-AL.
XX	XX
PD	25-FEB-1999.
XX	XX
PF	07-AUG-1998; 98WO-GB02383.
XX	XX
PR	13-AUG-1997; 97GB-0017192.
XX	XX
PA	(PLAN-) PLANT BIOSCIENCE LTD.
XX	XX
PI	Harberd NP, Peng J, Richards DE;
XX	XX
DR	WPI, 1999-181040/15.
XX	XX
PT	P-PSDB; AAT02540.
XX	XX
PT	New Triticum aestivum polynucleotides - encode a polypeptide which
PT	provides inhibition of the growth of plants, which inhibition is
PT	antagonised by gibberellin, used to confer a dwarf phenotype







[illegible]

PT provides inhibition of the growth of plants, which inhibition is  
 PT antagonised by gibberellin, used to confer a dwarf phenotype  
 XX  
 PS Disclosure: Flg 3a: 86pp; English.

CC The specification describes polypeptides encoded by the Rht gene (and  
CC its homologues) that, when expressed in *Triticum aestivum*, inhibit  
CC growth of the plant. This growth inhibition is antagonised by  
CC gibberellin. The products can be used to provide Rht expression in  
CC plants, conferring a dwarf phenotype on a plant which is correctable  
CC by treatment with gibberellin. In addition, the products can be  
CC used to produce Rht mutant plants which are dwarfed compared with  
CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants  
CC may be made by knocking out Rht or the relevant homologous gene in  
CC the plant of interest. Plants may be made which are resistant to  
CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol  
CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds  
CC dwarf but let crop plants grow tall. The present sequence represents  
CC the composite DNA sequence of wheat Rht gene.

SQ Sequence 2709 BP; 473 A; 900 C; 823 G; 482 T; 31 other;

Query Match	31.0%	Score 659	DB 20	Length 2709
Best Local Similarity	100.0%	Pred. No.	4.2e-274	
Matches 659; Conservative	0	Mismatches	0	Gaps 0

[illegible]

AA	AAX36285		ID AAX36285 standard; cDNA; 1746 BP.
XX			
AC	AAX36285;		
XX			
DT	16-JUL-1999	(first entry)	
XX			
DE	Consensus cDNA sequence of wheat Rht clone C15-1.		
XX			
KW	Rht gene; homologue; Triticum aestivum; wheat; growth inhibition; antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis; paclobutrazol; ss.		
XX			
OS	Triticum aestivum.		
PN	M0909174-A1.		
PD	25-FEB-1999.		
XX			
PF	07-AUG-1998; 98WO-GB02383.		
XX			
PR	13-AUG-1997; 97GB-0017192.		
XX			
PA	(PLAN-) PLANT BIOSCIENCE LTD.		
XX			
PI	Harberd NP, Peng J, Richards DE;		
DR	WPI; 1999-181040/15.		
XX			
PX	New Triticum Aestivum polynucleotides - encode a polypeptide which provides inhibition of the growth of plants, which inhibition is antagonised by gibberellin, used to confer a dwarf phenotype		
PS	Disclosure; Fig 2a; 8bpp; English.		
XX			
CC	The specification describes polypeptides encoded by the Rht gene (and its homologues) that, when expressed in Triticum Aestivum, inhibit growth of the plant. This growth inhibition is antagonised by gibberellin. The products can be used to provide Rht expression in plants, conferring a dwarf phenotype on a plant which is correctable by treatment with gibberellin. In addition, the products can be used to produce Rht mutant plants which are dwarfed compared with wild-type, the dwarfing being gibberellin-insensitive. Taller plants may be made by knocking out Rht or the relevant homologous gene in the plant of interest. Plants may be made which are resistant to compounds which inhibit gibberellin biosynthesis, such as paclobutrazol, e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds dwarf but let crop plants grow tall. The present sequence represents the consensus cDNA sequence of wheat Rht clone C15-1. Sequence 1746 BP; 375 A; 590 C; 495 G; 278 T; 8 other:		
SQ			
	Query Match	30.7%; Score 652; DB 20; Length 1746;	
	Best Local Similarity	100.0%; Pred. No. 4.5e-271;	
	Matches 652; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
OY	653 cggtcgccgcggcagcaacgcagcccgcgctcgctgctgctgcagcacgcagg	712	
DB	7 cggtcgccgcggcagcaacgcagcccgcgctcgctgctgctgcagcacgcagg	66	
OY	713 aggccggatttcgctgtgtgcacgcgcgtctggcgtgccccgaagccgtgcagcaggaga	772	
DB	67 aggccggatttcgctgtgtgcacgcgcgtctggcgtgccccgaagccgtgcagcaggaga	126	
OY	773 accttccgcgcgcggaaagcgcgtgtgtaaacagataaccttgctgagccgcgtaccagggcg	832	
DB	127 accttccgcgcgcggaaagcgcgtgtgtaaacagataaccttgctgagccgcgtaccagggcg	186	
OY	833 ggcgcgatgcgaagtgcgcgcactatcttcgcgcgaagccctgcgcgcgcgcgtcttcgcgt	892	
DB	187 ggcgcgatgcgaagtgcgcgcactatcttcgcgcgaagccctgcgcgcgcgcgtcttcgcgt	246	

QY	893	tccgcgcgaacgcggaacagctccctctctcgacgcgcctctgcgaactctctccacgcgc	952
Db	247	tccgcgcgcgaacgcgaacagctccctctctcgacgcgcctctgcgaactctctccacgcgc	306
QY	953	actcttagagtcctctgccccactcaagtctggacattcacgcgaacacagcattcc	1011
Db	307	actcttagagtcctctgccccactcaagtctggacattcacgcgaacacagcattcc	366
QY	1013	tggagagctctgcgcgcgcgcgtctgcacgtctgcattccgcattccacgcgaacagagga	1071
Db	367	tggagagctctgcgcgcgcgcgtctgcacgtctgcattccgcattccacgcgaacagagga	426
QY	1073	tgcagtgccgcgcacattctccagcctctgccttcgtcccgccgcgcctccctcttc	1131
Db	427	tgcagtgccgcgcacattctccagcctctgccttcgtcccgccgcgcctccctcttc	486
QY	1133	gctcaccgcgcgtctggccccccgcgaacgcgagacgcgaacgcctctgcacagtgcgt	1191
Db	487	gctcaccgcgcgtctggccccccgcgaacgcgagacgcgaacgcctctgcacagtgcgt	546
QY	1193	ggaagctgcgccagctctgcgcacacatccgcgtcgcattccagttaccgcgcctctgc	1251
Db	547	ggaagctgcgccagctctgcgcacacatccgcgtcgcattccagttaccgcgcctctgc	606
QY	1253	ccgcacgcgtccgcgagcctctgagccgttcatctgtcgcagccgagagcgagga	1304
Db	607	ccgcacgcgtccgcgagcctctgagccgttcatctgtcgcagccgagagcgagga	658
RESULT 5			
AAAX36268	AAAX36268 standard; DNA; 377 BP.		
AC	AAAX36268;		
DT	16-JUL-1999	(first entry)	
DE	DNA sequence obtained after sequencing wheat Rht clone 5a1.		
KM	Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;		
KW	antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;		
OS	Triticum aestivum.		
PN	WO9909174-A1.		
PD	25-FEB-1999.		
PF	07-AUG-1998; 98WO-G802383.		
PR	13-AUG-1997; 97GB-0017192.		
PA	(PLAN-) PLANT BIOSCIENCE LTD.		
PI	Harberd NP, Peng J, Richards DE;		
PI	WPI, 1999-181040/15.		
PT	New Triticum Aestivum polynucleotides - encode a polypeptide which		
PT	provides inhibition of the growth of plants, which inhibition is		
PT	antagonised by gibberellin, used to confer a dwarf phenotype		
PS	Disclosure: Fig 2c(1); 88pp; English.		
XX	The specification describes polypeptides encoded by the Rht gene (and		
XX	its homologues) that, when expressed in Triticum Aestivum, inhibit		
XX	growth of the plant. This growth inhibition is antagonised by		
XX	gibberellin. The products can be used to provide Rht expression in		
XX	plants, conferring a dwarf phenotype on a plant which is correctable		
XX	by treatment with gibberellin. In addition, the products can be		
XX	used to produce Rht mutant plants which are dwarfed compared with		
XX	wild-type, the dwarfing being gibberellin-insensitive. Taller plants		



DR MPI; 1999-181040/15.  
 DR P-PSDB; AAY02544.  
 XX  
 PT New Triticum aestivum polynucleotides - encode a polypeptide which  
 PT provides inhibition of the growth of plants, which inhibition is  
 XX antagonised by gibberellin, used to confer a dwarf phenotype  
 PS Disclosure; Fig 12a; 88pp; English.  
 XX  
 CC The specification describes polypeptides encoded by the Rht gene (and  
 CC its homologues) that, when expressed in Triticum aestivum, inhibit  
 CC growth of the plant. This growth inhibition is antagonised by  
 CC gibberellin. The products can be used to provide Rht expression in  
 CC plants, conferring a dwarf phenotype on a plant which is correctable  
 CC by treatment with gibberellin. In addition, the products can be  
 CC used to produce Rht mutant plants which are dwarfed compared with  
 CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants  
 CC may be made by knocking out Rht or the relevant homologue gene in  
 CC the plant of interest. Plants may be made which are resistant to  
 CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,  
 CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds  
 CC dwarf but let crop plants grow tall. The present sequence represents  
 CC the partial sequence of the wheat rht-10 allele.  
 CC  
 XX  
 SQ Sequence 416 BP; 59 A; 168 C; 129 G; 60 T; 0 other;

Query Match 10.2%; Score 217; DB 20; Length 416;  
 Best Local Similarity 100.0%; Pred. No. 6.7e-84;  
 Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 389 agtcaacgctcaccctcctcaccgctcagcgagcgagcgagcctctgactctccgcg 448  
 DB 200 agtcaacgctcaccctcctcaccgctcagcgagcgagcgagcctctgactctccgcg 259  
 OY 449 cctcgcctcgaactcctcagcagcatctacgcgctgcgcgcgacccctccgcgcgcg 508  
 DB 260 cctcgcctcgaactcctcagcagcatctacgcgctgcgcgcgacccctccgcgcgcg 319  
 OY 509 cgcagcgagcgagcgagcagcagctcgcgcgcgactcgctgcgcgcgaccccaagcgagtcgcgactg 568  
 DB 320 cgcagcgagcgagcgagcagcagctcgcgcgcgactcgctgcgcgcgaccccaagcgagtcgcgactg 379  
 OY 569 gcgggagcgagcagcagcagctcgcgcgcgactcgctgcgcgcgactcgctgcgcgcg 605  
 DB 380 gcgggagcgagcagcagcagctcgcgcgcgactcgctgcgcgcgactcgctgcgcgcg 416

RESULT 8  
 AAX36270  
 ID AAX36270 standard; DNA; 425 BP.  
 XX  
 AC AAX36270;  
 XX  
 DT 16-JUL-1999 (first entry)  
 XX  
 DE DNA sequence obtained after sequencing wheat Rht clone 5a1.  
 XX  
 KW Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;  
 KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;  
 KW paclobutrazol; ss.  
 XX  
 OS Triticum aestivum.  
 XX  
 PN WO9909174-A1.  
 XX  
 PD 25-FEB-1999.  
 XX  
 PF 07-AUG-1998; 98WO-GB02383.  
 XX  
 PR 13-AUG-1997; 97GB-0017192.  
 XX  
 PA (PLAN-) PLANT BIOSCIENCE LTD.

XX  
 PI Harberd NP, Peng J, Richards DE;  
 XX  
 DR MPI; 1999-181040/15.  
 XX  
 PT New Triticum aestivum polynucleotides - encode a polypeptide which  
 PT provides inhibition of the growth of plants, which inhibition is  
 XX antagonised by gibberellin, used to confer a dwarf phenotype  
 PS Disclosure; Fig 2c(3); 88pp; English.  
 XX  
 CC The specification describes polypeptides encoded by the Rht gene (and  
 CC its homologues) that, when expressed in Triticum aestivum, inhibit  
 CC growth of the plant. This growth inhibition is antagonised by  
 CC gibberellin. The products can be used to provide Rht expression in  
 CC plants, conferring a dwarf phenotype on a plant which is correctable  
 CC by treatment with gibberellin. In addition, the products can be  
 CC used to produce Rht mutant plants which are dwarfed compared with  
 CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants  
 CC may be made by knocking out Rht or the relevant homologue gene in  
 CC the plant of interest. Plants may be made which are resistant to  
 CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,  
 CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds  
 CC dwarf but let crop plants grow tall. The present sequence was  
 CC obtained after partially sequencing wheat Rht clone 5a1.  
 CC  
 XX  
 SQ Sequence 425 BP; 91 A; 125 C; 103 G; 94 T; 12 other;

Query Match 9.9%; Score 210; DB 20; Length 425;  
 Best Local Similarity 99.6%; Pred. No. 6.9e-81;  
 Matches 260; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1865 tcgcaacctgagatgagcgcttgcgcgcgagcgtgactcgcgagtttgaacgctgtaag 1924  
 DB 12 tcgcaacctgagatgagcgcttgcgcgcgagcgtgactcgcgagtttgaacgctgtaag 71  
 OY 1925 taacacatgagacatgaggaagaacaacagcccgagcgagcccgagcctccgagaa 1984  
 DB 72 taacacatgagacatgaggaagaacaacagcccgagcgagcccgagcctccgagaa 131  
 OY 1985 cgcagcgagcgagcagcagcttgaagaagaagctaaatgcatgctagtgagcgctgaat 2044  
 DB 132 cgcagcgagcgagcagcagcttgaagaagaagctaaatgcatgctagtgagcgctgaat 191  
 OY 2045 tgcagcgagcgagcagcagcttgcgcgcgactcgctgcgcgcgactcgctgcgcgcg 2104  
 DB 192 tgcagcgagcgagcagcagcttgcgcgcgactcgctgcgcgcgactcgctgcgcgcg 251  
 OY 2105 gctgagtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2125  
 DB 252 gctgagtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 272

RESULT 9  
 AAX36255  
 ID AAX36255 standard; DNA; 332 BP.  
 XX  
 AC AAX36255;  
 XX  
 DT 16-JUL-1999 (first entry)  
 XX  
 DE DNA sequence obtained after sequencing wheat Rht clone 14a1.  
 XX  
 KW Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;  
 KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;  
 KW paclobutrazol; ss.  
 XX  
 OS Triticum aestivum.  
 XX  
 PN WO9909174-A1.  
 XX  
 PD 25-FEB-1999.



```

XX (PLAN-) PLANT BIOSCIENCE LTD.
XX
XX Harberd NP, Peng J, Richards DE;
XX
XX WPI; 1999-181040/15.
XX
XX New Triticum aestivum polynucleotides - encode a polypeptide which
XX PT provides inhibition of the growth of plants, which inhibition is
XX PT antagonised by gibberellin, used to confer a dwarf phenotype
XX
XX PS Disclosure; Fig 2c(2); 88pp; English.
XX
XX The specification describes polypeptides encoded by the Rht gene (and
XX CC its homologues) that, when expressed in Triticum aestivum, inhibit
XX CC growth of the plant. This growth inhibition is antagonised by
XX CC gibberellin. The products can be used to provide Rht expression in
XX CC plants, conferring a dwarf phenotype on a plant which is correctable
XX CC by treatment with gibberellin. In addition, the products can be
XX CC used to produce Rht mutant plants which are dwarfed compared with
XX CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants
XX CC may be made by knocking out Rht or the relevant homologous gene in
XX CC the plant of interest. Plants may be made which are resistant to
XX CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,
XX CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds
XX CC dwarf but let crop plants grow tall. The present sequence was
XX CC obtained after partially sequencing wheat Rht clone 5a1.
XX
XX SQ Sequence 436 BP; 58 A; 153 C; 141 G; 66 T; 18 other;

Query Match          6.9%; Score 146; DB 20; Length 436;
Best Local Similarity 100.0%; Pred. No. 2.3e-53;
Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 454 gtcgactctccagcagcatctacgctgcgagccgagcccccgcggcgagcagc 513
    |||||||
DB 76 gtgactctccagcagcatctacgctgcgagccgagcccccgcggcgagcagc 135

QY 514 ggcgcggcgagcctgtcgcgcgagctcgtgcgggagcccaagcgagtcgagctggcggg 573
    |||||||
DB 136 ggcgcggcgagcctgtcgcgcgagctcgtgcgggagcccaagcgagtcgagctggcggg 195

QY 574 agcagcactcgtcgtcatctctctc 599
    |||||||
DB 196 agcagcactcgtcgtcatctctc 221

RESULT 12
AA36260
ID AAX36260 standard; DNA; 511 BP.
XX
XX AAX36260;
XX
XX 16-JUL-1999 (first entry)
XX
XX DNA sequence obtained after sequencing wheat Rht clone 14a1.
XX
XX Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;
XX KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;
XX KM paclobutrazol; ss.
XX
XX OS Triticum aestivum.
XX
XX PN WO9909174-A1.
XX
XX 25-FEB-1999.
XX
XX 07-AUG-1998; 98MO-GB02383.
XX
XX 13-AUG-1997; 97GB-0017192.
XX
XX (PLAN-) PLANT BIOSCIENCE LTD.
XX

```

```

XX Harberd NP, Peng J, Richards DE;
XX
XX WPI; 1999-181040/15.
XX
XX New Triticum aestivum polynucleotides - encode a polypeptide which
XX PT provides inhibition of the growth of plants, which inhibition is
XX PT antagonised by gibberellin, used to confer a dwarf phenotype
XX
XX PS Disclosure; Fig 2b(6); 88pp; English.
XX
XX The specification describes polypeptides encoded by the Rht gene (and
XX CC its homologues) that, when expressed in Triticum aestivum, inhibit
XX CC growth of the plant. This growth inhibition is antagonised by
XX CC gibberellin. The products can be used to provide Rht expression in
XX CC plants, conferring a dwarf phenotype on a plant which is correctable
XX CC by treatment with gibberellin. In addition, the products can be
XX CC used to produce Rht mutant plants which are dwarfed compared with
XX CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants
XX CC may be made by knocking out Rht or the relevant homologous gene in
XX CC the plant of interest. Plants may be made which are resistant to
XX CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,
XX CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds
XX CC dwarf but let crop plants grow tall. The present sequence was
XX CC obtained after partially sequencing wheat Rht clone 14a1.
XX
XX SQ Sequence 511 BP; 125 A; 137 C; 117 G; 119 T; 13 other;

Query Match          5.7%; Score 122; DB 20; Length 511;
Best Local Similarity 100.0%; Pred. No. 4.8e-43;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1883 gccctggccgggcccgtgattctcgcgagtttgaagcgtgtgaatcacatcgtgagcatgg 1942
    |||||||
DB 18 gcttgccgggcccgtgattctcgcgagtttgaagcgtgtgaatcacatcgtgagcatgg 77

QY 1943 aggaacaacagcccgcgagccgcgcgcgtctccggcgagcagcagcagcagcact 2002
    |||||||
DB 78 aggaacaacagcccgcgagccgcgcgcgtctccggcgagcagcagcagcagcagcact 137

QY 2003 tg 2004
    ||
DB 138 tg 139

RESULT 13
AA36259
ID AAX36259 standard; DNA; 357 BP.
XX
XX AAX36259;
XX
XX 16-JUL-1999 (first entry)
XX
XX DNA sequence obtained after sequencing wheat Rht clone 14a1.
XX
XX Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;
XX KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;
XX KM paclobutrazol; ss.
XX
XX OS Triticum aestivum.
XX
XX PN WO9909174-A1.
XX
XX 25-FEB-1999.
XX
XX 07-AUG-1998; 98MO-GB02383.
XX
XX 13-AUG-1997; 97GB-0017192.
XX
XX (PLAN-) PLANT BIOSCIENCE LTD.
XX
XX Harberd NP, Peng J, Richards DE;
XX

```

[illegible][illegible]



CC gibberellin. The products can be used to provide Rht expression in  
CC plants, conferring a dwarf phenotype on a plant which is correctable  
CC by treatment with gibberellin. In addition, the products can be  
CC used to produce Rht mutant plants which are dwarfed compared with  
CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants  
CC may be made by knocking out Rht or the relevant homologous gene in  
CC the plant of interest. Plants may be made which are resistant to  
CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol  
CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds  
CC dwarf but let crop plants grow tall. The present sequence was  
CC obtained after partially sequencing wheat Rht clone 14a1.  
XX  
50 Sequence 369 BP; 55 A; 142 C; 105 G; 53 T; 14 other:

	Query Match	4.2%	Score 89;	DB 20;	Length 369;	
	Best Local Similarity	99.3%;	Pred. No. 7.8e-29;			
	Matches 139;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;	
OY	. 1096	gcctcgcgacctcggtccaccggcgagcccttcctctcgttctcgcgctcaacggcgctcgcccccg	1155			
Dd	70	gccctcgcgctccgcgtcccgaggcgagcccttcctctcgttctcgcgctcaacggcgctcgcccccg	129			
OY	1156	cagccgcgacgagaaaccgagcccctgcagcagatgtggactgtgaagctctgcaccagtctcgac	1215			
Dd	130	cagccgcgacgagaaaccgagcccctgcagcagatgtggactgtgaagctctgcaccagtctcgac	189			
OY	1216	accatccgacgtctgacttcca	1235			
Dd	190	accatccgacgtctgacttcca	209			

Search completed: January 23, 2002, 04:58:07  
Job time: 7820 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 23, 2002, 02:36:51 ; Search time 1804.19 Seconds  
(without alignments)  
12656.527 Million cell updates/sec

Title: US-09-485-529-14  
Perfect score: 2125  
Sequence: 1 atagagagagagtagctc.....tgatgacgacgaactccg 2125

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 537289281 residues

Word size : 0

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:\*  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estlom:\*  
5: em\_estlpl:\*  
6: em\_estlba:\*  
7: em\_estrov:\*  
8: em\_estrov:\*  
9: em\_hlc:\*  
10: gb\_estl:\*  
11: gb\_est2:\*  
12: gb\_hlc:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_trod:\*  
20: em\_gss\_vrl:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	87	4.1	400	10	BE414891 MML001.H0
2	77	3.6	268	11	BE201178 WHE0986_A
3	77	3.6	597	11	BE620181 HVSMEC001
4	68	3.2	778	11	BE308975 HVSMEC000
5	51	2.4	551	10	AT770444 606052P10
6	51	2.4	844	11	BE268018 HV_CEA001
7	46	2.2	559	11	BE618807 HVSMEC000
8	45	2.1	475	10	BE364489 P11_14_F0
9	45	2.1	481	11	BE417150 949053G08
10	45	2.1	527	10	BE595338 P11_48_G1
11	45	2.1	563	11	BE588097 FMI_37_F0
12	44	2.1	592	11	BE713881 EMI_60_F1

13	44	2.1	658	11	BE587541 FMI_37_F0
14	44	2.1	977	10	BE196286 HVSMEC009
15	44	2.1	981	10	BE454560 HVSMEC000
16	41	1.9	851	11	BE254247 HVSMEC000
17	38	1.8	440	11	BE109897 IPI_36_A1
18	35	1.6	228	10	AU095344 AU095344
19	35	1.6	263	11	D39460 RICS0803A R
20	35	1.6	388	11	C27475 C27475 RICE
21	35	1.6	399	10	AU091413 AU091413
22	35	1.6	416	11	BE051785 FMI_58_D0
23	35	1.6	443	11	BE655617 FMI_46_D1
24	35	1.6	467	10	AU222715 AU222715
25	35	1.6	596	10	AU065169 AU065169
26	29	1.4	259	11	BE411689 BE411689
27	29	1.4	427	10	AI577799 UT-R-ABO-
28	26	1.2	700	10	AL506960 AL506960
29	25	1.2	593	10	AI881894 606074G08
30	24	1.1	373	10	AI967572 LjItrppest
31	24	1.1	417	10	AV407563 AV407563
32	24	1.1	420	10	AV409142 AV409142
33	24	1.1	453	10	AM464360 BP230015B
34	24	1.1	566	10	AM066993 683012H10
35	24	1.1	586	11	BE419686 LjNEST47e
36	23	1.1	223	10	AU009349 AU009349
37	23	1.1	223	10	AU009350 AU009350
38	23	1.1	283	10	AU007007 AU007007
39	23	1.1	283	10	AU007010 AU007010
40	23	1.1	353	10	AU006981 AU006981
41	23	1.1	358	10	AU008620 AU008620
42	23	1.1	358	10	AU008657 AU008657
43	23	1.1	360	10	AU013337 AU013337
44	23	1.1	361	11	D24385 RRCR1819A R
45	23	1.1	362	10	AU009468 AU009468

#### ALIGNMENTS

RESULT 1  
BE414891  
LOCUS MML001.H03P90430 ITFC MML Wheat Root Library Triticum aestivum  
DEFINITION CDNA clone MML001.H03, mRNA sequence.  
ACCESSION BE414891.1 GI:9412639  
VERSION BE414891  
KEYWORDS EST.  
SOURCE Triticum aestivum  
ORGANISM bread wheat.

REFERENCE  
AUTHORS Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P., Herrmann,R.G., Holton,T., Jacquemin,I.M., Jia,T., Joudrier,P., Langridge,P., Lazo,G.R., Lin,D.J., McGuire,P., Ogihara,T., Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M., Sorrells,M., Warburton,M. and Wenzel,G.  
International Triticaceae EST Cooperative (ITFC): Production of  
Expressed Sequence Tags for Species of the Triticaceae  
Unpublished (2000)  
Contact: Warburton M  
Applied Biotechnology Center, CIMMYT  
Apdo. Postal 6-641, 06600 Mexico DF MEXICO  
Tel: 52-5-7269091 ext 1381  
Fax: 52-5-7267558/59  
Email: mwarburton@cgnet.com  
International Triticaceae EST Cooperative (ITFC)  
http://wheat.pw.usda.gov/genome.  
Location/Qualifiers  
1. 400  
/organism="Triticum aestivum"  
/cultivar="Atlas"

TITLE  
JOURNAL  
COMMENT

FEATURES  
source

BASE COUNT	52 a	165 c	109 g	67 t	7 others
ORIGIN					
Query Match	4.1%;	Score 87;	DB 10;	Length 400;	
Best Local Similarity	100.0%;	Pred. No. 5.8e-27;			
Matches 87;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
Oy 999	caaccaggcatcctgtgagcgcttcgccgctgcgcgcgcgctgcaagctgcgttgcactgg 1058				
Db 112	CAACACGACGCAATCCGTCGAGGCGGTTCGCCGCGCTGCCGCGCGTCGACGTCGATGTCGG 171				
Oy 1059	catcagcagggggatgcagtgcccg 1085				
Db 172	CATCAGCAGGGGATGATGACGCGCCG 198				
RESULT 2					
LOCUS	BF201178	268 bp	mRNA	EST	06-NOV-2000
DEFINITION	WHE0986_A04_B0825 Wheat pre-anthesis spike cDNA library Triticum aestivum cDNA clone WHE0986_A04_B08, mRNA sequence.				
ACCESSION	BF201178				
VERSION	BF201178.1	GI:1115968			
KEYWORDS	EST.				
SOURCE	bread wheat.				
ORGANISM	Triticum aestivum				
REFERENCE	Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han, P.S., Hala, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J., Seaton, C.L. and Tong, J.C.				
AUTHORS	The structure and function of the expressed portion of the wheat genomes - Pre-anthesis spike cDNA library				
JOURNAL	Unpublished (2000)				
COMMENT	Contact: Olin Anderson US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center 800 Buchanan Street, Albany, CA 94710, USA Tel: 5105595773 Fax: 5105595818 Email: oanderson@pw.usda.gov Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20 Seq primer: Stragene SK primer.				
FEATURES	Location/Qualifiers				
SOURCE	1..268				
	/organism="Triticum aestivum"				
	/cultivar="Chinese Spring"				
	/db_xref="taxon:4565"				
	/clone="WHE0986_A04_B08"				
	/clone_lib="wheat pre-anthesis spike cDNA library"				
	/tissue_type="Spike before anthesis"				
	/dev_stage="Adult plant"				
	/lab_host="E. coli SOLR"				
	/note="Vector: Lambda Uni-ZAP XR, excised phagemid; site_1: EcoRI; site_2: XhoI; Plants were grown in the greenhouse. Whole spike with awns trimmed, white, green and yellow anther were collected and total RNA, and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give plasmid phagemids in the RJ Close lab (Choi, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD				

```

BASE COUNT      40 a       97 c       84 g       47 t
ORIGIN
Anderson Lab (all other authors).

Query Match          3.6%; Score 77; DB 11; Length 268;
Best Local Similarity 100.0%; Pred. No. 1,1e-22;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY   1345    gtcttcgagatgcaccggcgtctcgccgacgccggcgcccttgagaaggctcttggcac 1404
Db     18    GTCTTCGAGATGCACCGGGCTGTCCGCCAGCCCCGCCCTTGAGAAGTCTCTGGCAC 77
              |||
OY   1405    gtgcggccgctgtggccc 1421
              |||
Db     78    GTGC GGCGCCTGTGGGCC 94
              |||

RESULT 3
BF620181
LOCUS DEFINITION BF620181 597 bp mRNA EST 22-FEB-2001
HVSMEC0018M1lf Hordeum vulgare seedling shoot EST library
HVCDNA0003 (Etisolated and unstrressed) Hordeum vulgare CDNA clone
HVSMEC0018M1lf, mRNA sequence.
BF620181
ACCESSION BF620181.2 GI:13109232
VERSION EST.
KEYWORDS Hordeum vulgare
SOURCE barley.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophytes; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Hordeum.
I (bases 1 to 597)
Wing,R., Close,T.J., Kleinhoofs,A., Wise,R., Begun,D., Frisch,D., Yu
,Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo
,T., Saski,C., Schwartzbeck,U., Simmons,J., Chol.D.W., Main,D. and
Wood,T.
Development of a genetically and physically anchored EST resource
for barley genomics
Unpublished (2000)
On Dec 18, 2000 this sequence version replaced gi:11883915.
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29654, USA
Tel.: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: AATTACCCTCATCAAAAGG
High quality sequence stop: 564.
Location/Qualifiers
        1..597
         /organism="Hordeum vulgare"
         /cultivar="Morex"
         /db_xref="taxon:4513"
         /clone="HVSMEC0018M1lf"
         /clone_lip="hordeum vulgare seedling shoot EST library
HVCDNA0003 (Etisolated and unstrressed)"
         /tisue_type="Seedling shoot"
         /lab_host="TJUC121"
         /note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI; For
more details on library preparation and sequence analysis
see http://www.genome.clemson.edu/projects/barley/ To
order a clone see http://www.genome.clemson.edu/orders"

BASE COUNT      99 a       202 c       204 g       91 t
ORIGIN

```

```

DB      288  GCGCGCGCGGACGACGAGTCATGTCGAGGTACTCGCGCGGAGATCTCAGAC 347
              |||
QY      1648  gtgtgtgctctgcgaggg 1664
              |||
DB      348  GTGTGTGCTGCTGCGAGGG 364
              |||

RESULT  4
BG308975  778 bp  mRNA  EST  22-FEB-2001
LOCUS     HG308975
DEFINITION HVSMC0001C18f Hordeum vulgare seedling shoot EST library
            HVCNMA0003 (Etiolated and unstressed) Hordeum vulgare cDNA clone
            HVSMC0001C18f, mRNA sequence.
ACCESSION BG308975
VERSION   BG308975.1 GI:13109822
KEYWORDS  EST.
SOURCE    Hordeum vulgare
ORGANISM  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
            ; Triticeae; Hordeum.
            1 (bases 1 to 778)
            Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Fritsch,D., Yu
            ,Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo
            ,T., Saski,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D. and
            Wood,T.
            Development of a genetically and physically anchored EST resource
            for barley genomics
            Unpublished (2000)
JOURNAL   Contact: Wing RA
COMMENT   Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: AATTACCTCTCACTAAGG
            High quality sequence stop: 690.
            Location/Qualifiers
            source
              1..778
                /organism="Hordeum vulgare"
                /cultivar="Morex"
                /db_xref="taxon:4513"
                /clone="HVSMEC0001C18f"
                /clone_lib="Hordeum vulgare seedling shoot EST library
                HVCNMA0003 (Etiolated and unstressed)"
                /tissue_type="Seedling shoot"
                /lab_host="TJc121"
                /note="Vector: lambdaZAP, Site_1: EcoRI, Site_2: XhoI; For
                more details on library preparation and sequence analysis
                see http://www.genome.clemson.edu/projects/barley/ To
                order a clone see http://www.genome.clemson.edu/orders"
BASE COUNT 96 a 295 c 261 g 124 t 2 others

Query Match 3.2%; Score 68; DB 11; Length 778;
Best Local Similarity 100.0%; Pred. No. 9.3e-19;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

DEFINITION 606052F10.x2 606 - Ear tissue cDNA library from Schmidt lab Zea
            mays cDNA, mRNA sequence.
ACCESSION A1770444
VERSION   A1770444.1 GI:5268480
KEYWORDS  EST.
SOURCE    Zea mays.
ORGANISM  Zea mays.
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoidae; Andropogoneae; Zea.
            1 (bases 1 to 551)
            Walbot,V.
            Maize ESTs from various cDNA libraries sequenced at Stanford
            University
            Unpublished (1999)
JOURNAL   Contact: Walbot V
COMMENT   Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 606052 row: F column: 10.
            Location/Qualifiers
            source
              1..551
                /organism="Zea mays"
                /cultivar="Ohio43"
                /db_xref="taxon:4577"
                /clone_lib="606 - Ear tissue cDNA library from Schmidt
                lab"
                /tissue_type="mixed"
                /dev_stage="ear length from 0.5 cm - 2.0 cm"
                /lab_host="XJOLR (Stratagene)"
                /note="Organ: Immature ear; Vector: pBK-CMV; Site_1: EcoRI
                ; Site_2: XhoI; Mixed ear tissue cDNA library from Schmidt
                lab"
BASE COUNT 66 a 213 c 184 g 88 t

Query Match 2.4%; Score 51; DB 10; Length 551;
Best Local Similarity 100.0%; Pred. No. 1.9e-11;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Seq primer: AATTAACCTCCTAAGG  
High quality sequence stop: 587.  
Location/Qualifiers

## FEATURES

source

1..844

/organism="Hordeum vulgare"  
/cultivar="C116155 (M1a13)"  
/db\_xref="taxon:4513"

/clone="HV\_CEA0019j17f"

/clone.lib="Hordeum vulgare seedling green leaf EST  
library HVCDNA0004 (Erysiphe infected & control)"  
/tissue.type="seedling green leaf"

/note="Vector: lambdaZAP; Site\_1: EcoRI; Site\_2: XhoI; For  
more details on library preparation and sequence analysis  
see http://www.genome.clemson.edu/projects/barley/ To  
order a clone see http://www.genome.clemson.edu/orders"

BASE COUNT 150 a 290 c 276 g 128 t

ORIGIN

Query Match 2.4%; Score 51; DB 11; Length 844;  
Best Local Similarity 100.0%; Pred. No. 2e-11;  
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 178 tccgacatgacgacgtgacgacgaagctgacgacgtcgaagatggccatg 228  
|||||  
Db 299 TCCGACATGCGGACGTGGCGAAGAGCTGAGACGCTGACATGGCCATG 349

RESULT 7

BF618807

LOCUS BF618807 559 bp mRNA EST 22-FEB-2001  
DEFINITION HVMEC0007P09f Hordeum vulgare seedling shoot EST library  
HVCDNA0003 (Etiolated and unstressed) Hordeum vulgare cDNA clone  
HVMEC0007P09f, mRNA sequence.

ACCESSION BF618807

VERSION BF618807.1 GI:1182541

KEYWORDS EST.

SOURCE barley.

ORGANISM Hordeum vulgare

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
; Triticeae; Hordeum.

Wing,R., Close,T.J., Kleinholz,A., Wise,R., Begum,D., Fritsch,D., Yu  
,Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo  
,T., Saski,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Kain,D. and  
Wood,T.

Development of a genetically and physically anchored EST resource  
for barley genomes  
Unpublished (2000)

TITLE

JOURNAL

COMMENT

Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Seq primer: AATTAACCTCCTAAGG  
High quality sequence start: 2  
High quality sequence stop: 405.  
Location/Qualifiers

FEATURES

source

1..559

/organism="Hordeum vulgare"  
/cultivar="Morex"

/db\_xref="taxon:4513"  
/clone="HVMEC0007P09f"

/clone.lib="Hordeum vulgare seedling shoot EST library  
HVCDNA0003 (Etiolated and unstressed)"

/tissue.type="Seedling shoot"

/lab\_host="TJUC121"

/note="Vector: lambdaZAP; Site\_1: EcoRI; Site\_2: XhoI; For  
more details on library preparation and sequence analysis  
see http://www.genome.clemson.edu/projects/barley/ To  
order a clone see http://www.genome.clemson.edu/orders"

BASE COUNT 100 a 172 c 197 g 90 t

ORIGIN

Query Match 2.2%; Score 46; DB 11; Length 559;  
Best Local Similarity 100.0%; Pred. No. 2.7e-09;  
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1259 ccctgcggacctgagaccgttcacatgctcagccgagggcgagga 1304  
|||||  
Db 1 CGCTCGCGGACCTGAGCGCTTCATGCTGACCGAGGCGAGGA 46

RESULT 8

BE364489

LOCUS BE364489 475 bp mRNA EST 20-JUL-2000  
DEFINITION P11\_14\_F02.b1.A002 Pathogen induced 1 (P11) Sorghum bicolor cDNA,  
mRNA sequence.

ACCESSION BE364489

VERSION BE364489.1 GI:9306046

KEYWORDS EST.

SOURCE Sorghum.

ORGANISM Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoidae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 475)

Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt  
,L.H.

An EST database from Sorghum: pathogen-induced plants  
Unpublished (2000)

CONTACT: Cordonnier-Pratt MM

JOURNAL

COMMENT

The University of Georgia  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 542 1805

Email: mmp@prattuga.edu

Sequences have been trimmed to exclude polyA, vector and regions  
below Phred quality 16. The threshold for highest quality sequence  
is 20.

Seq primer: JEN REV

High quality sequence stop: 469

POLYA-No.

Location/Qualifiers

FEATURES

source

1..475

/organism="Sorghum bicolor"  
/db\_xref="taxon:4558"

/clone.lib="Pathogen induced 1 (P11)"

/note="Organ: Anthracnose-infected leaves from  
two-week-old sorghum plants 48 hr after inoculation;  
Vector: pBluescript II from Lambda zap II; Site\_1: XhoI;  
Site\_2: EcoRI; Two-week-old sorghum plants (B7x 623  
cultivar) were infected with pathogen (isolate FRM421 of  
Colletotrichum graminicola, which is a sorghum isolate).

RNA was prepared from infected leaves harvested from 45  
seedlings 48 hours after inoculation. Note: young  
seedlings (2 weeks old) exhibit juvenile resistant  
reaction, which is an incompatible interaction. As they  
grow older (4 weeks or older), plants resume susceptibility  
to anthracnose disease. The library was made from poly-A  
RNA in the cloning vector lambda zap II. Clones to be  
sequenced were prepared by mass excision. WARNING: While  
most or all ESTs are expected to derive from the host  
plant, no effort was made to eliminate ESTs deriving from



DEFINITION	FM1_37_F08.bl_A003 Floral-Induced Meristem 1 (FMI) Sorghum propinquum cDNA, mRNA sequence.					
ACCESSION	BF588097					
VERSION	BF588097.1 GI:11680421					
KEYWORDS	EST.					
SOURCE	Sorghum propinquum.					
ORGANISM	Sorghum propinquum Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum. 1 (bases 1 to 563) Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt, L.H.					
REFERENCE						
AUTHORS	An EST database from Sorghum: floral-induced meristems Unpublished (2000)					
TITLE	JOURNAL					
COMMENT	Contact: Cordonnier-Pratt MM Department of Botany The University of Georgia Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel.: 706 542 1860 Fax: 706 542 1805 Email: mmprratt@uga.edu					
FEATURES	Seq primer: JEN REV High quality sequence stop: 477 POLYA-No.					
SOURCE	Location/Qualifiers 1..563 /organism="Sorghum propinquum" /db_xref="taxon:132711" /clone_id="Floral-Induced Meristem 1 (FMI)" /note="Organ: Floral-induced meristems; Vector: pBluescript II from lambda Zap II; Site.1: XhoI; Site.2: EcoRI; mature plants were placed in a growth chamber for 15 days with 16 hr darkness and 8 hr light (flowering is induced by short-day conditions); 16 days after being returned to the greenhouse under natural long days during late April/early May, meristems were harvested The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."					
BASE COUNT	76 a 217 c 175 g 95 t					
ORIGIN						
Query Match	2.1% Score 45; DB 11; Length 563;					
Best Local Similarity	100.0%; Pred. No. 7.4e-09;					
Matches 45; Conservative	0; Mismatches 0; Indels 0; Gaps 0.					
OY	730 gtgcacgagcgtctgcgtcgatgcggaggccgtgcagcaagaac 774       Db 67 GTGCACGCCGCTCCTCGCTGCCGGAGCCGTGCACGACGAC 111					
RESULT 12						
LOCUS	BG713881 592 bp mRNA EST 08-MAY-2001					
DEFINITION	EM1_60_F10.g2_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA sequence.					
ACCESSION	BG713881					
VERSION	BG713881.1 GI:14007831					
KEYWORDS	EST.					
SOURCE	Sorghum.					
ORGANISM	Sorghum bicolor Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum. 1 (bases 1 to 592) Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H. An EST database from Sorghum: developing embryos Unpublished (2000)					

COMMENT	Contact: Cordonnier-Pratt MM Department of Botany The University of Georgia Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 542 1805 Email: mmpratt@uga.edu Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
FEATURES	Seq primer: Polymix High quality sequence start: 9 High quality sequence stop: 563 POLYA-No.  Location/Qualifiers 1..592 /organism="Sorghum bicolor" /db_xref="taxon:4558" /clone_lib="Embryo 1 (EM1)" /note="Organ: Embryos germinated for 24 hr. Vector: plasmidcript II from lambda Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the clonin vector lambda Zap II. Clones to be sequenced were prepared by mass excision."
BASE COUNT	99 a       179 c       184 g       130 t
ORIGIN	
Query Match	2.1%; Score 44; DB 11; Length 592;
Best Local Similarity	100.0%; Pred. No. 2e-08;
Matches 44; Conservative	0; Mismatches 0; Indels 0; Gaps 0
QY 1741	cacctggcctccaatgctctacaagcagcgacgcgtgctgac 1784       DB 107 CACCTGGGCTCAATGCCTACAAGCAGCGACGACGTGTGCC 150
RESULT 13	
LOCUS	BF587541 658 bp mRNA EST 12-DEC-2000
DEFINITION	F01.37.F08.g1.A003 floral-induced Meristem 1 (F01) Sorghum
VERSION	propinquum cDNA, mRNA sequence.
ACCESSION	BF587541
KEYWORDS	EST.
SOURCE	Sorghum propinquum.
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum. 1 (bases 1 to 658) Cordonnier-Pratt,M.-W., Gingle,A., Sudman,M., Marsala,C. and Pratt, L.H.
REFERENCE	An EST database from Sorghum: floral-induced meristems
AUTHORS	Unpublished (2000)
TITLE	Contact: Cordonnier-Pratt MM
JOURNAL	Department of Botany
COMMENT	The University of Georgia Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 542 1805 Email: mmpratt@uga.edu Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
FEATURES	Seq primer: Polymix High quality sequence start: 35 High quality sequence stop: 658 POLYA-No.  Location/Qualifiers 1..658 /organism="Sorghum propinquum" /db_xref="taxon:132711"
FEATURES	
SOURCE	



/clone\_11b="Floral-Induced Meristem 1 (FM1)"  
 /note="Organ: Floral-Induced meristems; Vector:  
 pBluescript II from Lambda Zap II; Site.1: XhoI; Site.2:  
 EcoRI; mature plants were placed in a growth chamber for  
 15 days with 16 hr darkness and 8 hr light (flowering is  
 induced by short-day conditions); 16 days after being  
 returned to the greenhouse under natural long days during  
 late April/early May, meristems were harvested. The  
 library was made from poly-A RNA in the cloning vector  
 lambda Zap II. Clones to be sequenced were prepared by  
 mass excision."

BASE COUNT 115 a 196 c 205 g 142 t  
 ORIGIN

Query Match 2.1%; Score 44; DB 11; Length 658;  
 Best Local Similarity 100.0%; Pred. No. 2e-08;  
 Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1741 caccctggctccaatgcctacaagcagcagcagctgtgc 1784  
 ||||||||||||||||||||||||||||||||||||||||  
 Db 144 CACCTGGCTCCAAATGCTACAGCAGCGACGACCTGCTGCG 187

RESULT 14  
 BE196286 977 bp mRNA EST 02-MAR-2001  
 LOCUS HVSMH0091004f Hordeum vulgare 5-45 DAP spike EST library  
 DEFINITION HVCNMA0009 (5 to 45 DAP) Hordeum vulgare cDNA clone HVSMH0091004f,  
 mRNA sequence.

ACCESSION BE196286  
 VERSION BE196286  
 KEYWORDS

SOURCE  
 ORGANISM  
 Hordeum vulgare  
 barley.

REFERENCE  
 AUTHORS  
 Wing,R., Close,T.J., Kleinofs,A., Wise,R., Begum,D., Frisch,D., Yu  
 ,Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo  
 ,T., Saski,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D. and  
 Wood,T.  
 Development of a genetically and physically anchored EST resource  
 for barley genomics  
 Unpublished (2000)  
 On Jun 26, 2000 this sequence version replaced gi:8708481.

JOURNAL  
 COMMENT  
 Contact: Wing RA  
 Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 7288  
 Fax: 864 656 4293  
 Email: twing@clemson.edu  
 Seg primer: AATTACCTCCTCACTAAAGG  
 High quality sequence stop: 618.

FEATURES  
 Source

1..977  
 /organism="Hordeum vulgare"  
 /cultivar="Morex"  
 /db\_xref="taxon:4513"  
 /clone="HVSMH0091004f"  
 /clone\_11b="Hordeum vulgare 5-45 DAP spike EST library  
 HVCNMA0009 (5 to 45 DAP)"  
 /tissue\_type="5-45 DAP Spike"  
 /lab\_host="SOLR"  
 /note="Vector: LambdaZAP; Site.1: EcoRI; Site.2: XhoI; For  
 more details on library preparation and sequence analysis  
 see http://www.genome.clemson.edu/projects/barley/ To  
 order a clone see http://www.genome.clemson.edu/orders"

BASE COUNT 234 a 272 c 270 g 197 t 4 others  
 ORIGIN

Query Match 2.1%; Score 44; DB 10; Length 977;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-08;  
 Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1741 caccctggctccaatgcctacaagcagcagcagctgtgc 1784  
 ||||||||||||||||||||||||||||||||||||||||  
 Db 37 CACCTGGCTCCAAATGCTACAGCAGCGACGACGCTGCTGCG 80

RESULT 15  
 BE454560 981 bp mRNA EST 02-MAR-2001  
 LOCUS HVSMH0094E04f Hordeum vulgare 5-45 DAP spike EST library  
 DEFINITION HVCNMA0009 (5 to 45 DAP) Hordeum vulgare cDNA clone HVSMH0094E04f,  
 mRNA sequence.

ACCESSION BE454560  
 VERSION BE454560  
 KEYWORDS

SOURCE  
 ORGANISM  
 Hordeum vulgare  
 barley.

REFERENCE  
 AUTHORS  
 Wing,R., Close,T.J., Kleinofs,A., Wise,R., Begum,D., Frisch,D., Yu  
 ,Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo  
 ,T., Saski,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D. and  
 Wood,T.  
 Development of a genetically and physically anchored EST resource  
 for barley genomics  
 Unpublished (2000)  
 On Jul 26, 2000 this sequence version replaced gi:9463094.

JOURNAL  
 COMMENT  
 Contact: Wing RA  
 Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 7288  
 Fax: 864 656 4293  
 Email: twing@clemson.edu  
 Seg primer: AATTACCTCCTCACTAAAGG  
 High quality sequence stop: 672.  
 Location/Qualifiers

FEATURES  
 source

1..981  
 /organism="Hordeum vulgare"  
 /cultivar="Morex"  
 /db\_xref="taxon:4513"  
 /clone="HVSMH0094E04f"  
 /clone\_11b="Hordeum vulgare 5-45 DAP spike EST library  
 HVCNMA0009 (5 to 45 DAP)"  
 /tissue\_type="5-45 DAP Spike"  
 /lab\_host="SOLR"  
 /note="Vector: LambdaZAP; Site.1: EcoRI; Site.2: XhoI; For  
 more details on library preparation and sequence analysis  
 see http://www.genome.clemson.edu/projects/barley/ To  
 order a clone see http://www.genome.clemson.edu/orders"

BASE COUNT 277 a 244 c 241 g 190 t 29 others  
 ORIGIN

Query Match 2.1%; Score 44; DB 10; Length 981;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-08;  
 Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1741 caccctggctccaatgcctacaagcagcagcagctgtgc 1784  
 ||||||||||||||||||||||||||||||||||||||||  
 Db 18 CACCTGGCTCCAAATGCTACAGCAGCGACGACGCTGCTGCG 61

Search completed: January 23, 2002., 04:03:30  
 Job time: 5199 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 23, 2002, 03:34:27 ; Search time 78.5 Seconds  
(without alignments)  
6130.774 Million cell updates/sec

Title: US-09-485-529-14

Perfect score: 2125

Sequence: 1 atagagagggcaggtagctc.....tgatgtgacgacgactcgcg 2125

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 11323899 residues

Word size : 0

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents.NA:\*  
1: /cgn2\_6/prodata/2/lna/5A.COMB.seq:\*  
2: /cgn2\_6/prodata/2/lna/5B.COMB.seq:\*  
3: /cgn2\_6/prodata/2/lna/6A.COMB.seq:\*  
4: /cgn2\_6/prodata/2/lna/6B.COMB.seq:\*  
5: /cgn2\_6/prodata/2/lna/PCRTOS.COMB.seq:\*  
6: /cgn2\_6/prodata/2/lna/Backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	1.0	4308	3	US-08-938-291A-8
2	21	1.0	1491	4	US-09-082-092-9
3	20	0.9	590	1	US-08-314-309A-10
4	20	0.9	1028	4	US-08-118-200-1
5	20	0.9	1028	4	US-08-458-745-1
6	20	0.9	1461	1	US-08-403-634-1
7	20	0.9	1461	4	US-08-913-441B-1
8	20	0.9	2036	3	US-08-923-454A-17
9	20	0.9	2205	3	US-08-888-077A-41
10	20	0.9	3172	1	US-08-314-309A-1
11	20	0.9	3517	2	US-08-642-406A-20
12	20	0.9	3517	3	US-08-434-000A-1
13	20	0.9	3517	4	US-09-312-157-1
14	20	0.9	9997	1	US-08-246-982A-15
15	20	0.9	9997	1	US-08-453-265-15
16	20	0.9	10103	2	US-08-457-273B-7
17	20	0.9	10348	2	US-08-457-273B-41
18	20	0.9	10348	3	US-08-556-419-13
19	20	0.9	10348	4	US-09-041-886-14
20	20	0.9	10366	1	US-08-246-982A-5
21	20	0.9	10366	1	US-08-453-265-5
22	19	0.9	71	2	US-08-712-948-4
23	19	0.9	472	2	US-08-465-380-33
24	19	0.9	472	2	US-08-486-397-33
25	19	0.9	472	2	US-08-486-399-33
26	19	0.9	472	2	US-08-461-965-33
27	19	0.9	472	2	US-08-634-641-33

c 28	19	0.9	472	3	US-09-249-471-33	Sequence 33, Appl
c 29	19	0.9	472	3	US-09-249-472-33	Sequence 33, Appl
c 30	19	0.9	472	3	US-09-249-451-33	Sequence 33, Appl
c 31	19	0.9	472	3	US-08-809-453-33	Sequence 33, Appl
c 32	19	0.9	472	3	US-09-249-461-33	Sequence 33, Appl
c 33	19	0.9	472	3	US-09-249-448-33	Sequence 33, Appl
c 34	19	0.9	472	2	US-08-465-380-35	Sequence 35, Appl
c 35	19	0.9	472	2	US-08-486-397-35	Sequence 35, Appl
c 36	19	0.9	477	2	US-08-486-399-35	Sequence 35, Appl
c 37	19	0.9	477	2	US-08-461-965-35	Sequence 35, Appl
c 38	19	0.9	477	2	US-08-634-641-35	Sequence 35, Appl
c 39	19	0.9	477	3	US-09-249-471-35	Sequence 35, Appl
c 40	19	0.9	477	3	US-09-249-472-35	Sequence 35, Appl
c 41	19	0.9	477	3	US-09-249-451-35	Sequence 35, Appl
c 42	19	0.9	477	3	US-08-809-455-35	Sequence 35, Appl
c 43	19	0.9	477	3	US-09-249-461-35	Sequence 35, Appl
c 44	19	0.9	477	3	US-09-249-448-35	Sequence 35, Appl
c 45	19	0.9	3358	2	US-08-469-537A-104	Sequence 104, App

## ALIGNMENTS

RESULT 1  
US-08-938-291A-8/c  
; Sequence 8, Application US/08938291A  
; Patent No. 6117673  
; GENERAL INFORMATION:  
; APPLICANT: Lev, Sima  
; APPLICANT: Plowman, Gregory D.  
; APPLICANT: Schlessinger, Joseph  
; TITLE OF INVENTION: RDB PROTEINS AND RELATED  
; TITLE OF INVENTION: PRODUCTS AND METHODS  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; STREET: Suite 4700  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071-2066  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
; MEDIUM TYPE: storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: FastSeq  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/938,291A  
; FILING DATE: September 26, 1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/027,337  
; FILING DATE: October 11, 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard J.  
; REGISTRATION NUMBER: 32,327  
; REFERENCE/DOCKET NUMBER: 228/172  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELETYPE: 67-3510  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4308 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-938-291A-8  
Query Match 1.0%; Score 22; DB 3; Length 4308;



APPLICANT: NAGARAJA, Ramalah  
APPLICANT: KREMER, Eric J  
APPLICANT: YU, Sui  
APPLICANT: BAKER, Elizabeth  
APPLICANT: MULLEY, John C  
APPLICANT: MANDEL, Jean-Louis  
APPLICANT: PRITCHARD, Melanie April  
APPLICANT: LYNCH, Michael  
TITLE OF INVENTION: DNA SEQUENCES RELATED TO ISOLATED  
NUMBER OF SEQUENCES: 13  
TITLE OF INVENTION: FRAGILE X SYNDROME  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/118,200  
FILING DATE: 09-SEP-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/802,650  
FILING DATE: 05-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/672,232  
FILING DATE: 20-MAR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/638,518  
FILING DATE: 04-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/966,517  
FILING DATE: 23-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feury, Sharon E  
REGISTRATION NUMBER: 36,113  
REFERENCE/DOCKET NUMBER: 020160-164  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1028 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-118-200-1

Query Match 0.9%; Score 20; DB 4; Length 1028;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 59 gcggcgagcgagtgagcgc 78  
|||||

Db 396 GCGGCGGCGGCGGTGCGGC 415

RESULT 5  
US-08-458-745-1  
; Sequence 1, Application US/08458745  
; Patent No. 6242576  
; GENERAL INFORMATION:  
; APPLICANT: SUTHERLAND, Grant R  
; APPLICANT: RICHARDS, Robert I  
; APPLICANT: SCHLESSINGER, David

APPLICANT: NAGARAJA, Ramalah  
APPLICANT: KREMER, Eric J  
APPLICANT: YU, Sui  
APPLICANT: BAKER, Elizabeth  
APPLICANT: MULLEY, John C  
APPLICANT: MANDEL, Jean-Louis  
APPLICANT: PRITCHARD, Melanie April  
APPLICANT: LYNCH, Michael  
TITLE OF INVENTION: DNA SEQUENCES RELATED TO ISOLATED  
NUMBER OF SEQUENCES: 13  
TITLE OF INVENTION: FRAGILE X SYNDROME  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/458,745  
FILING DATE: 02-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/118,200  
FILING DATE: 09-SEP-1993  
APPLICATION NUMBER: US 07/802,650  
FILING DATE: 05-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/672,232  
FILING DATE: 20-MAR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/638,518  
FILING DATE: 04-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/966,517  
FILING DATE: 23-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feury, Sharon E  
REGISTRATION NUMBER: 36,113  
REFERENCE/DOCKET NUMBER: 020160-164  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1028 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-458-745-1

Query Match 0.9%; Score 20; DB 4; Length 1028;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 59 gcggcgagcgagtgagcgc 78  
|||||

Db 396 GCGGCGGCGGCGGTGCGGC 415

RESULT 6  
US-08-403-634-1/c  
; Sequence 1, Application US/08403634  
; Patent No. 5674748  
; GENERAL INFORMATION:  
; APPLICANT: Giordano, Antonio

```

; TITLE OF INVENTION: NOVEL HUMAN CYCLIN-DEPENDENT
; TITLE OF INVENTION: KINASE-LIKE PROTEINS AND METHODS
; TITLE OF INVENTION: OF USING THE SAME
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Woodcock, Washburn, Kurtz, Mackiewicz &
; ADDRESSER: No. 56747815
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,634
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/208,575
; FILING DATE: 08-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-1482
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1461 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 65..1180
; US-08-403-634-1

Query Match      0.9%; Score 20; DB 1; Length 1461;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 354 caacgcgcgcgcgcgcgcgc 373
Db 55 CAACGCCGCCGCCGCCGCC 36

RESULT 7
US-08-913-441B-1/c
; Sequence 1, Application US/08913441B
; Patent No. 6162612
; GENERAL INFORMATION:
; APPLICANT: Giordano, Antonio
; TITLE OF INVENTION: No. 6162612el Human Cyclin-Dependent Kinase-Like Proteins and
; TITLE OF INVENTION: Methods of Using The Same
; FILE REFERENCE: 8321-76 C11
; CURRENT APPLICATION NUMBER: US/08/913,441B
; CURRENT FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 08/403,634
; PRIOR FILING DATE: 1995-03-14
; PRIOR APPLICATION NUMBER: PCT/US96/03557
; PRIOR FILING DATE: 1996-03-14
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1461
; TYPE: DNA
```

```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (65)..(1180)
; US-08-913-441B-1

Query Match      0.9%; Score 20; DB 4; Length 1461;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 354 caacgcgcgcgcgcgcgcgc 373
Db 55 CAACGCCGCCGCCGCCGCC 36

RESULT 8
US-08-923-454A-17/c
; Sequence 17, Application US/08923454A
; Patent No. 6004794
; GENERAL INFORMATION:
; APPLICANT: Creasy, Caretha
; APPLICANT: Livi, George
; APPLICANT: Kairan, Eric
; APPLICANT: Clinkenbeard, Helen
; APPLICANT: Browne, Michael
; APPLICANT: Southan, Christopher
; TITLE OF INVENTION: HUMAN SERINE PROTEASE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSER: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923,454A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/025436
; FILING DATE: 06-SEPT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Baumeister, Kirk
; REGISTRATION NUMBER: 33,833
; REFERENCE/DOCKET NUMBER: P50547
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5096
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2036 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; MOLECULE TYPE: CDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORIGINAL SOURCE: Feature polymorphism at 1325
; US-08-923-454A-17

Query Match      0.9%; Score 20; DB 3; Length 2036;
Best Local Similarity 100.0%; Pred. No. 11;
```

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 ggcggcgagcggtgagcg 78  
|||||  
Db 38 GCGCGCGCGCGGTGGCGC 19

## RESULT 9

US-08-888-077A-41/C  
; Sequence 41, Application US/08888077A  
; Patent No. 6020143  
; GENERAL INFORMATION:  
; APPLICANT: ST. GEORGE-HYSLOP, PETER H  
; APPLICANT: ROMMENS, JOHANNA M  
; APPLICANT: FRASER, PAUL E  
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED  
; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE AND USDS THEREFOR.  
; NUMBER OF SEQUENCES: 41  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LERNER, DAVID, LITTENBERG, KROMHOLZ & MENTLIK  
; STREET: 600 SOUTH AVENUE WEST  
; CITY: WESTFIELD  
; STATE: NJ  
; COUNTRY: USA  
; ZIP: 07090-1497  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/888,077A  
; FILING DATE: 03-JUL-1997  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/592,541  
; FILING DATE: 26-JAN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PALISI, THOMAS M  
; REGISTRATION NUMBER: 36,629  
; REFERENCE/DOCKET NUMBER: SCHERING 3.0-017 CIP CIP CIP IV  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (908) 654-5000  
; TELEFAX: (908) 654-7866  
; INFORMATION FOR SEQ ID NO: 41:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2205 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: misc-feature  
; LOCATION: 1..2205  
; OTHER INFORMATION: /note="mutm1-TM2"  
US-08-888-077A-41

Query Match 0.9%; Score 20; DB 3; Length 2205;  
Best local Similarity 100.0%; Pred. No. 11;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 ggcggcgagcggtgagcg 78  
|||||  
Db 113 GCGCGCGCGCGGTGGCGC 94

## RESULT 10

US-08-314-309A-1/C  
; Sequence 1, Application US/08314309A  
; Patent No. 5677141  
; GENERAL INFORMATION:  
; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: FUKAGAWA, MASAO

APPLICANT: IWAMI, MORITA  
APPLICANT: ARAMORI, ICHIRO  
APPLICANT: KOJO, HITOSHI  
TITLE OF INVENTION: PROCESS FOR PRODUCING 7-AMINOCEPHEM  
TITLE OF INVENTION: COMPOUND OR SALTS THEREOF  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,  
ADDRESS: P.C.  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/314,309A  
FILING DATE: 30-SEP-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/631,906  
FILING DATE: 21-DEC-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: OBLON, No. 5677141man F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 18-863-0 CONT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEX: 248655 OPAT UR  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3172 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
US-08-314-309A-1

Query Match 0.9%; Score 20; DB 1; Length 3172;  
Best local Similarity 100.0%; Pred. No. 11;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 ggcggcgagcggtgagcg 142  
|||||  
Db 484 GGAGGGGAGGAGGTGACG 465

## RESULT 11

US-08-642-406A-20/C  
; Sequence 20, Application US/08642406A  
; Patent No. 5959177  
; GENERAL INFORMATION:  
; APPLICANT: Hein, Mich B.  
; APPLICANT: Hatt, Andrew C.  
; APPLICANT: Ma, Julian K.C.  
; TITLE OF INVENTION: TRANSGENIC PLANTS EXPRESSING ASSEMBLED  
; TITLE OF INVENTION: SECRETORY ANTIBODIES  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: THE SCRIPPS RESEARCH INSTITUTE  
; STREET: 10666 No. 5959177th Torrey Pines Road, TPC-8  
; CITY: La Jolla  
; STATE: California  
; COUNTRY: US  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/642,406A
FILING DATE: 03-MAY-1996
CLASSIFICATION: 800

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/591,823
FILING DATE: 02-OCT-1990

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US
FILING DATE: 27-OCT-1989

ATTORNEY/AGENT INFORMATION:
NAME: Logan, April C.
REGISTRATION NUMBER: 33,950
REFERENCE/DOCKET NUMBER: 184.2

TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 554-2937
TELEFAX: (619) 554-6312

INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 3517 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: gDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO

```

Query Match	0.98;	Score 20;	DB 2;	Length 3517;
Best Local Similarity	100.0%;	Pred. No. 11;		
Matches 20;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

QY 59 gcgcgcgcgcgcgtgcgcgc 78  
|||  
Db 2486 GCGCGCGCGCGGTGCGGC 2467

Db 2486 GCGGCGGCGGCGGTGGCGGC 2467

RESULT 12  
 US-08-434-000A-1/C  
 Sequence 1, Application US/08434000A  
 Patent No. 6046037  
 GENERAL INFORMATION:  
 APPLICANT: ANDREW C. HIATT, JULIAN  
 APPLICANT: K.-C. MA, THOMAS LEHNER  
 TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION  
 TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES  
 NUMBER OF SEQUENCES: 19  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Lyon & Lyon  
 STREET: 633 West Fifth Street  
 STREET: Suite 4700  
 CITY: Los Angeles  
 STATE: California  
 COUNTRY: U.S.A.  
 ZIP: 90071  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
 MEDIUM TYPE: Storage  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: IBM P.C. DOS 5.0  
 SOFTWARE: Word Perfect 5.1  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/434,000A  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 PRIOR APPLICATION DATA: including application  
 PRIOR APPLICATION DATA: described below:  
 APPLICATION NUMBER: 08/367,395

```

1      FILING DATE: 12/30/94
2      ATTORNEY/AGENT INFORMATION:
3      NAME: Guise, Jeffrey W.
4      REGISTRATION NUMBER: 34,613
5      REFERENCE/DOCKET NUMBER: 212/127
6      TELECOMMUNICATION INFORMATION:
7      TELEPHONE: (619) 552-8400
8      TELEFAX: (619) 552-0159
9      TELEX: 67-3510
10     TELETYPE:
11     INFORMATION FOR SRO ID NO: 1:
12     SEQUENCE CHARACTERISTICS:
13     LENGTH: 3517 base pairs
14     TYPE: nucleic acid
15     STRANDEDNESS: single
16     TOPOLOGY: linear
17     FEATURE:
18     NAME/KEY: Coding Sequence
19     LOCATION: 124.....2445
20     OS-08-434-000A-1

```

Query Match 0.9%; Score 20; DB 3; Length 3517;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 20; Conservative 0; Mismatches 0; Gaps 0;

QY	59	gcgcgcgcgcgtgcgc	78
Db	2486	gcgcgcgcgcgtgcgc	2467

Db 2486 GCGGCGGCGGCGGTGGCGGC 2467

RESULT 13  
 US-09-312-157-1/c  
 : Sequence 1, Application US/09312157  
 : Patent NO. 6303341  
 :  
 : GENERAL INFORMATION:  
 : APPLICANT: ANDREW C. HIATT, JULIAN  
 : K.-C. MA, THOMAS LEHNER  
 : TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION  
 : IMMUNOGLOBULINS IN PLANTS AND THEIR USES  
 :  
 : NUMBER OF SEQUENCES: 19  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Lyon & Lyon  
 : STREET: 633 West Fifth Street  
 : Suite 4700  
 : CITY: Los Angeles  
 : STATE: California  
 : COUNTRY: U.S.A.  
 : ZIP: 90071  
 :  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
 : storage  
 :  
 : COMPUTER: IBM Compatible  
 : OPERATING SYSTEM: IBM P.C. DOS 5.0  
 : SOFTWARE: Word Perfect 5.1  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/09/312,157  
 : FILING DATE: 14-May-1999  
 : CLASSIFICATION: <Unknown>  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: 08/434,000  
 : FILING DATE: <Unknown>  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Guise, Jeffrey W.  
 : REGISTRATION NUMBER: 34,613  
 : REFERENCE/DOCKET NUMBER: 212/127  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: (619) 552-8400  
 : TELEFAX: (619) 552-0159  
 : TELEX: 67-351  
 :  
 : SEQUENCE LISTING  
 : INFORMATION FOR SEQ ID NO: 1



```

:      SEQUENCE CHARACTERISTICS:
:      LENGTH: 3517 base pairs
:      TYPE: nucleic acid
:      STRANDEDNESS: single
:      TOPOLOGY: linear
:
:      FEATURE:
:      NAME/KEY: Coding Sequence
:      LOCATION: 124...2445
:      SEQUENCE DESCRIPTION: SEQ ID NO: 1
:
US-09-312-157-1

```

	Query Match	0.9%;	Score 20;	DB 4;	Length 3517;
	Best Local Similarity	100.0%	Pred. No. 11;		
Matches	20;	Conservative	0;	Mismatches	0;
				Indels	Gaps
Qy	59	gcggcgcgcgcgctgtgcgcgc	78		
Db	2486	gcgcgcgcgcgcgcgcgcgcgc	2467		

```

RESULT 14
US-08-246-982A-15/c
Sequence 15 Application US/08246982A
Patent No. 5686288
GENERAL INFORMATION:
APPLICANT: Macdonald, Marcy E.
APPLICANT: Ambrose, Christine M.
APPLICANT: Duyao, Mabel P.
APPLICANT: Gussella, James F.
TITLE OF INVENTION: Huntingtin DNA, Protein And Uses Thereof
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/246,982A
FILING DATE: May 20, 1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Goldstein, Jorge, A.
REGISTRATION NUMBER: 29,021
REFERENCE/DOCKET NUMBER: 0609.3880002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 9997 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 90..9446
US-08-246-982A-15

```

	Query Match	0.9%	Score 20	DB 1	Length 9977
	Best Local Similarity	100.0%	Pred. No. 9.5		
	Matches	20	Conservative	0	Mismatches 0; Indels 0; Gaps 0.
OY	59 gcgcgcgcgcgcgcgttgccgac	78			

Db 195 GCGGCGGCGGCGGTGGCGC 176

RESULT 15  
US-08-453-265-15/c  
: Sequence 15, Application US/08453265  
: Patent No. 5693757  
: GENERAL INFORMATION:  
: APPLICANT: Macdonald, Marcy E.  
: APPLICANT: Ambrose, Christine M.  
: APPLICANT: Duyao, Mabel P.  
: APPLICANT: Gusella, James F.  
: TITLE OF INVENTION: Huntington DNA, Protein And Uses Thereof  
: NUMBER OF SEQUENCES: 25  
: CORRESPONDENCE ADDRESSES:  
: ADDRESSEE: Stetene, Kessler, Goldstein & Fox  
: STREET: 1100 New York Avenue  
: CITY: Washington  
: STATE: D.C.  
: COUNTRY: U.S.A.  
: ZIP: 20005  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/453,265  
: FILING DATE: 30-May-1995  
: CLASSIFICATION: 514  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Ludwig, Steven R.  
: REGISTRATION NUMBER: 36,203  
: REFERENCE/DOCKET NUMBER: 0609.3880003  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (202) 371-2600  
: TELEFAX: (202) 371-2540  
: INFORMATION FOR SEQ ID NO: 15:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 9997 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: both  
: TOPOLOGY: linear  
: FEATURE:  
: NAME/KEY: CDS  
: LOCATION: 90..9446  
: US-08-453-265-15

	Query Match	Best Local Similarity	0.9%; 100.0%;	Score 20; Pred. No. 9.5;	DB 1;	Length 997;
Matches	20;	Conservative	0;	Mismatches	0;	Indels
Gaps	0;					
OY	59	gcgcgcgcgcgcgcgtcgcgcgc	78			
db	195	gcgcgcgcgcgcgcgcctgacgcgc	176			

Search completed: January 23, 2002, 05:00:55  
Job time: 5192 sec

\_\_\_\_\_



XX PS Disclosure; Fig 8b; 88bp; English.  
CC CC The specification describes polypeptides encoded by the Rht gene (and  
CC its homologues) that, when expressed in *Triticum aestivum*, inhibit  
CC growth of the plant. This growth inhibition is antagonised by  
CC gibberellin. The products can be used to provide Rht expression in  
CC plants, conferring a dwarf phenotype on a plant which is correctable  
CC by treatment with gibberellin. In addition, the products can be  
CC used to produce Rht mutant plants which are dwarfed compared with  
CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants  
CC may be made by knocking out Rht or the relevant homologous gene in  
CC the plant of interest. Plants may be made which are resistant to  
CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,  
CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds  
CC dwarf but let crop plants grow tall. The present sequence is encoded by  
CC the wheat Rht clone 5a1 genomic sequence.  
XX SQ Sequence 623 AA;

Query Match 100.0%; Score 13; DB 20; Length 623;  
Best Local Similarity 100.0%; Pred. NO. 0.00044;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LNAPPPPLPPAPQ 13  
Db 109 LNAPPPPLPPAPQ 121  
|||||

RESULT 2  
AAG27119  
ID AAG27119 standard; Protein; 33 AA.  
XX AC AAG27119;  
XX DT 17-OCT-2000 (first entry)  
XX DE Zea mays protein fragment SEQ ID NO: 31833.  
XX KW Protein identification; signal transduction pathway; metabolic pathway;  
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX KW termination sequence; corn.  
XX OS Zea mays subsp. mays.  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-0301439.  
XX PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139730.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142053.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 27-JUL-1999; 99US-0145919.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 03-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.

PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 22-OCT-1999; 99US-0161004.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 29-OCT-1999; 99US-0162142.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 69.2%; Score 9; DB 21; Length 33;  
 Best Local Similarity 100.0%; Pred. No. 0.11;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLPPAP 12  
 Db 13 PPPPLPPAP 21

RESULT 3  
 AAB80893 standard; protein; 182 AA.  
 XX  
 AC AAB80893;  
 XX  
 DT 30-MAY-2001 (first entry)  
 XX  
 DE Human ATFX leucine zipper domain.  
 XX  
 KW Human; ATFX; gamma aminobutyric acid B receptor; GABA B receptor;  
 KM transcription factor; leucine zipper.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200116596-A2.  
 XX  
 PD 08-MAR-2001.  
 XX  
 PF 31-AUG-2000; 2000WO-GB03361.  
 XX  
 PR 31-AUG-1999; 99GB-0020569.  
 PR 12-JAN-2000; 2000GB-0000516.  
 XX  
 PA (GLAX ) GLAXO GROUP LTD.  
 XX  
 PI White J, Wise A, Marshall F;  
 XX  
 DR WPI; 2001-265904/27.  
 DR N-PSDB; AAF77878.  
 XX  
 PT Identifying modulators of gamma aminobutyric acid(GABA)-B receptor  
 PT mediated activity by monitoring the interaction between GABAB receptor  
 PT and the CREB/ATF transcription factors in the presence of a test  
 PT substance  
 XX  
 PS Claim 14; Page 64-65; 71pp; English.  
 XX  
 CC The present invention relates to a method for identifying a modulator of  
 CC gamma aminobutyric acid\_B (GABA\_B) receptor-mediated activity; by  
 CC monitoring the interaction between a CREB/ATF transcription factor  
 CC capable of binding to GABA\_B receptor. The present sequence is the  
 CC leucine zipper domain of human ATFX (see AAB80892), which was used in the  
 CC method of the present invention. ATFX is a member of the CREB/ATF family  
 CC of transcription factors. ATFX contains a bZIP domain, by which it  
 CC interacts with the coiled coil domain of GABA\_B receptor. Modulators of  
 CC GABA\_B receptor activity are useful for treating central nervous system  
 CC or peripheral nervous system disorders.  
 XX  
 SQ Sequence 182 AA;

Query Match 69.2%; Score 9; DB 22; Length 182;  
 Best Local Similarity 100.0%; Pred. No. 0.45;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPAP 12  
 |||||  
 DB 31 PPPPLPPAP 39

RESULT 4  
 AAB80892  
 ID AAB80892 standard; protein: 216 AA.

AC AAB80892;  
 XX  
 DT 30-MAY-2001 (first entry)

DE Human ATFX.

KW Human: ATFX; gamma aminobutyric acid B receptor; GABA B receptor;  
 KW transcription factor.

OS Homo sapiens.

PN WO200116596-A2.

PD 08-MAR-2001.

XX 31-AUG-2000; 2000WO-GB03361.

XX 31-AUG-1999; 99GB-0020569.

PR 12-JAN-2000; 2000GB-0000516.

XX (GLAX ) GLAXO GROUP LTD.

PI White J, Wise A, Marshall F;

DR MPI: 2001-265904/27.

XX N-PSDB; AAF77877.

XX Identifying modulators of gamma aminobutyric acid(GABA)-B receptor  
 PT mediated activity by monitoring the interaction between GABAB receptor  
 PT and the CREB/ATF transcription factors in the presence of a test  
 PT substance

PS Claim 14; Fig 1; 71pp; English.

XX The present invention relates to a method for identifying a modulator of  
 CC gamma aminobutyric acid,B (GABA,B) receptor-mediated activity, by  
 CC monitoring the interaction between a CREB/ATF transcription factor  
 CC capable of binding to GABA,B receptor. The present invention is human  
 CC ATFX, which was used in the method of the present invention. ATFX is a  
 CC member of the CREB/ATF family of transcription factors. ATFX contains a  
 CC bZIP domain, by which it interacts with the coiled coil domain of GABA,B  
 CC receptor. Modulators of GABA,B receptor activity are useful for treating  
 CC central nervous system or peripheral nervous system disorders.

XX Sequence 216 AA:

Query Match 69.2%; Score 9; DB 22; Length 216;  
 Best Local Similarity 100.0%; Pred. No. 0.51;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPAP 12  
 |||||  
 DB 65 PPPPLPPAP 73

RESULT 5  
 AAW21733

ID AAW21733 standard; Protein: 223 AA.

XX AAW21733;

XX 01-OCT-1997 (first entry)

DE NIP-1 encoded by clone 59.

KW NIP-1; NIP-2; NuMA; nuclear mitotic apparatus; NuMA interacting protein;  
 KW cell division; proliferation; antibody; Ab; detection;  
 KW malignant cell growth.

OS Homo sapiens.

PN WO9640917-A1.

PD 19-DEC-1996.

XX 07-JUN-1996; 96WO-US09504.

PR 07-JUN-1995; 95US-0478408.

PA (UYUA ) UNIV YALE.

PI McPherson SMG, Snyder MP;

DR MPI: 1997-077270/07.

XX N-PSDB; AAT77784.

XX New nucleic acid encoding nuclear mitotic appts. interacting  
 PT proteins - useful for modulating cell division and proliferation and  
 PT in diagnosis  
 PS Claim 28; Page 57-58; 78pp; English.

XX The sequences given in AAW21733-35 represent NuMA binding proteins (NuMA  
 CC = nuclear mitotic apparatus). These protein sequences were identified  
 CC using the fusion proteins given in AAW21731-32. Compounds which  
 CC interfere with the interaction of NuMA with a known NIP (NuMA interacting  
 CC protein) are used to modulate cell division and/or proliferation. Ab,  
 CC raised conventionally using NIP-1 or -2 as immunogen, are used to detect  
 CC NIP (or their complexes) and to block their activity for diagnostic or  
 CC therapeutic use, e.g. to detect defective NuMA or NIP which may be  
 CC markers for aberrant (including malignant) cell growth (which can also  
 CC be detected by nucleic acid sequencing). Also where malignancy is  
 CC related to defects in NuMA or NIP, it can be treated by administration  
 CC of the appropriate functional protein. This protein is rich in  
 CC proline residues.

XX Sequence 223 AA:

Query Match 69.2%; Score 9; DB 18; Length 223;  
 Best Local Similarity 100.0%; Pred. No. 0.53;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPAP 12  
 |||||  
 DB 2 PPPPLPPAP 10

RESULT 6

AAB95859  
 ID AAB95859 standard; Protein: 282 AA.

XX AAB95859;

AC 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:18922.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

OS Homo sapiens.  
 XX  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-0116126.  
 XX  
 PR 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI; 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 PS Claim 8; SEQ ID 18922; 2537pp + CD ROM; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX  
 SQ Sequence 282 AA;

Query Match 69.2%; Score 9; DB 22; Length 282;  
 Best Local Similarity 100.0%; Pred No. 0.64;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLPPAP 12  
 |||||||||  
 Db 131 PPPPLPPAP 139

RESULT 7  
 AAB80896  
 ID AAB80896 standard; protein; 282 AA.  
 XX  
 AC AAB80896;  
 XX  
 DT 30-MAY-2001 (first entry)  
 XX  
 DE Human ATRX with N-terminal extension.  
 XX

KM Human: ATRX; gamma aminobutyric acid B receptor; GABA B receptor;  
 KM transcription factor.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200116596-A2.  
 XX  
 PD 08-MAR-2001.  
 XX  
 PF 31-AUG-2000; 2000WO-GB03361.  
 XX  
 PR 31-AUG-1999; 99GB-0020569.  
 PR 12-JAN-2000; 2000GB-0000516.  
 XX  
 PA (GLAX ) GLAXO GROUP LTD.  
 XX  
 PI White J, Wise A, Marshall F;  
 PI N-PSDB; AAF77880.  
 XX  
 DR WPI; 2001-265904/27.  
 XX  
 PT Identifying modulators of gamma aminobutyric acid(GABA)-B receptor  
 PT mediated activity by monitoring the interaction between GABAB receptor  
 PT and the CREB/ATF transcription factors in the presence of a test  
 PT substance -  
 XX  
 PS Claim 14; Fig 9; 71pp; English.  
 XX  
 CC The present invention relates to a method for identifying a modulator of  
 CC gamma aminobutyric acid<sub>B</sub> (GABA<sub>B</sub>) receptor-mediated activity, by  
 CC monitoring the interaction between a CREB/ATF transcription factor  
 CC capable of binding to GABA<sub>B</sub> receptor. The present sequence is human  
 CC ATRX, which was used in the method of the present invention. ATRX is a  
 CC member of the CREB/ATF family of transcription factors. ATRX contains a  
 CC bZIP domain, by which it interacts with the coiled coil domain of GABA<sub>B</sub>  
 CC receptor. Modulators of GABA<sub>B</sub> receptor activity are useful for treating  
 CC central nervous system or peripheral nervous system disorders. The  
 CC present sequence has an N-terminal extension compared to AAB80892.  
 XX  
 SQ Sequence 282 AA;

Query Match 69.2%; Score 9; DB 22; Length 282;  
 Best Local Similarity 100.0%; Pred No. 0.64;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLPPAP 12  
 |||||||||  
 Db 131 PPPPLPPAP 139

RESULT 8  
 AAG07315  
 ID AAG07315 standard; Protein; 183 AA.  
 XX  
 AC AAG07315;  
 XX  
 DT 17-OCT-2000 (first entry)  
 XX  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 4422.  
 XX  
 KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 PN EP1033405-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 25-FEB-2000; 2000EP-0301439.  
 XX

PR	25-FEB-1999;	99US-0121825;
PR	05-MAR-1999;	99US-0123180;
PR	09-MAR-1999;	99US-0123548;
PR	23-MAR-1999;	99US-0125788;
PR	25-MAR-1999;	99US-0126264;
PR	29-MAR-1999;	99US-0126785;
PR	01-APR-1999;	99US-0127462;
PR	06-APR-1999;	99US-0128233;
PR	08-APR-1999;	99US-0128714;
PR	16-APR-1999;	99US-0129845;
PR	19-APR-1999;	99US-0130077;
PR	21-APR-1999;	99US-0130449;
PR	23-APR-1999;	99US-0130510;
PR	23-APR-1999;	99US-0130891;
PR	28-APR-1999;	99US-0131449;
PR	30-APR-1999;	99US-0132048;
PR	30-APR-1999;	99US-0132407;
PR	04-MAY-1999;	99US-0132484;
PR	05-MAY-1999;	99US-0132485;
PR	06-MAY-1999;	99US-0132486;
PR	06-MAY-1999;	99US-0132867;
PR	07-MAY-1999;	99US-0132863;
PR	11-MAY-1999;	99US-0134256;
PR	14-MAY-1999;	99US-0134218;
PR	14-MAY-1999;	99US-0134219;
PR	14-MAY-1999;	99US-0134321;
PR	14-MAY-1999;	99US-0134370;
PR	18-MAY-1999;	99US-0134768;
PR	19-MAY-1999;	99US-0134941;
PR	20-MAY-1999;	99US-0135124;
PR	21-MAY-1999;	99US-0135353;
PR	24-MAY-1999;	99US-0135629;
PR	25-MAY-1999;	99US-0136021;
PR	27-MAY-1999;	99US-0136392;
PR	28-MAY-1999;	99US-0136782;
PR	01-JUN-1999;	99US-0137222;
PR	03-JUN-1999;	99US-0137528;
PR	04-JUN-1999;	99US-0137502;
PR	07-JUN-1999;	99US-0137724;
PR	08-JUN-1999;	99US-0138094;
PR	10-JUN-1999;	99US-0138540;
PR	10-JUN-1999;	99US-0138847;
PR	14-JUN-1999;	99US-0139114;
PR	16-JUN-1999;	99US-0139452;
PR	16-JUN-1999;	99US-0139453;
PR	17-JUN-1999;	99US-0139492;
PR	18-JUN-1999;	99US-0139454;
PR	18-JUN-1999;	99US-0139455;
PR	18-JUN-1999;	99US-0139460;
PR	18-JUN-1999;	99US-0139461;
PR	18-JUN-1999;	99US-0139462;
PR	18-JUN-1999;	99US-0139463;
PR	18-JUN-1999;	99US-0139750;
PR	18-JUN-1999;	99US-0139763;
PR	21-JUN-1999;	99US-0139817;
PR	22-JUN-1999;	99US-0139899;
PR	23-JUN-1999;	99US-0140353;
PR	23-JUN-1999;	99US-0140354;
PR	24-JUN-1999;	99US-0140695;
PR	28-JUN-1999;	99US-0140823;
PR	29-JUN-1999;	99US-0140991;
PR	30-JUN-1999;	99US-0141287;
PR	01-JUL-1999;	99US-0141842;
PR	01-JUL-1999;	99US-0142154;
PR	02-JUL-1999;	99US-0142055;
PR	06-JUL-1999;	99US-0142390;
PR	08-JUL-1999;	99US-0142803;
PR	09-JUL-1999;	99US-0142920;
PR	12-JUL-1999;	99US-0142927;
PR	25-FEB-1999;	99US-0143342;
PR	05-MAR-1999;	99US-0143622;
PR	09-MAR-1999;	99US-0144005;
PR	23-MAR-1999;	99US-0144085;
PR	25-MAR-1999;	99US-0144086;
PR	29-MAR-1999;	99US-0144325;
PR	01-APR-1999;	99US-0144331;
PR	06-APR-1999;	99US-0144332;
PR	08-APR-1999;	99US-0144334;
PR	16-APR-1999;	99US-0144335;
PR	19-APR-1999;	99US-0144632;
PR	20-JUL-1999;	99US-0144844;
PR	21-JUL-1999;	99US-0144814;
PR	21-JUL-1999;	99US-0145086;
PR	21-JUL-1999;	99US-0145088;
PR	22-JUL-1999;	99US-0145085;
PR	22-JUL-1999;	99US-0145087;
PR	22-JUL-1999;	99US-0145089;
PR	23-JUL-1999;	99US-0145192;
PR	23-JUL-1999;	99US-0145145;
PR	23-JUL-1999;	99US-0145224;
PR	26-JUL-1999;	99US-0145276;
PR	27-JUL-1999;	99US-0145913;
PR	27-JUL-1999;	99US-0145918;
PR	27-JUL-1999;	99US-0145919;
PR	28-JUL-1999;	99US-0145951;
PR	02-AUG-1999;	99US-0146386;
PR	02-AUG-1999;	99US-0146388;
PR	02-AUG-1999;	99US-0146589;
PR	03-AUG-1999;	99US-0147038;
PR	04-AUG-1999;	99US-0147204;
PR	05-AUG-1999;	99US-0147302;
PR	05-AUG-1999;	99US-0147192;



PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 61.5%; Score 8; DB 21; Length 183;  
Best Local Similarity 100.0%; Pred. No. 3.3;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPA 11  
|||||||  
Db 167 PPPPLPPA 174

## RESULT 9

AAG07314 standard; Protein; 188 AA.

AC AAG07314;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 4421.

KW Protein identification; signal transduction pathway; metabolic pathway;  
hybridisation assay; genetic mapping; gene expression control; promoter;  
termination sequence.

KW Arabidopsis thaliana.

OS Arabidopsis thaliana.

PN EPI03405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 30-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144332.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145224.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 28-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156456.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.

PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160747.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 61.5%; Score 8; DB 21; Length 188;  
Best Local Similarity 100.0%; Pred. No. 3.3;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPA 11  
|||||  
Db 172 PPPPLPPA 179

RESULT 10  
AAG07313  
ID AAG07313 standard; Protein: 195 AA.  
XX  
AC AAG07313;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SPQ ID NO: 4420.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hydridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
OS Arabidopsis thaliana.  
XX  
PM EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135153.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137724.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138647.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142820.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145089.  
PR 23-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146389.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151338.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.

PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 61.5%; Score 8; DB 21; Length 195;  
Best Local Similarity 100.0%; Pred. No. 3.4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPA 11  
|||||||  
DB 179 PPPPLPPA 186

RESULT 11  
AAB32786  
ID AAB32786 standard; Protein: 234 AA.

XX AC AAB32786;

XX DT 25-JAN-2001 (first entry)

XX DE Eucalyptus grandis transcription factor protein sequence #244.

XX KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;  
KW poplar; sweetgum; teak; mangany; bZIP; G-box binding factor;  
KW basic helix-loop-helix zipper; homeotic; homeobox; MADS;  
KW homeodomain zipper; LIM domain; AP2; EREBS; zinc finger domain;  
KW type 2 Cys2His2; CCAAT box element; MYB.

XX OS Eucalyptus grandis.

XX PN WO200053724-A2.

XX PD 14-SEP-2000.

XX PF 09-MAR-2000; 2000WO-US06112.

XX PR 11-MAR-1999; 99US-0266513.

XX PR 18-AUG-1999; 99US-0149485.

XX PA (GENE-) GENESIS RES & DEV CORP LTD.

XX PI (FLET-) FLETCHER CHALLENGE FORESTS LTD.

XX DR Wood M, McGrath A, Shenk MA, Glenn M;

XX WI WI: 2000-579369/54.

XX PT New isolated polynucleotide encoding a plant transcription factor for

XX PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,

XX PT having modified gene expression or modified activity of a polypeptide

XX PS Claim 8; Page 304; 747pp; English.

XX CC The present invention relates to novel plant transcription factors from

CC Eucalyptus grandis or Pinus radiata. The present sequence is one such

CC transcription factor. The transcription factor may be used to produce a  
CC plant having modified gene expression such as a woody plant e.g. a  
CC eucalyptus, pine, acacia, poplar, sweetgum, teak, or mangany species or  
CC to modify the activity of a polypeptide in a plant. The transcription  
CC factors of the present invention are members from the following families  
CC of regulatory proteins: bZIP, bZIP family of G-box binding factors, basic  
CC helix-loop-helix zipper, homeotic/homeodomain/homeobox/MADS, homeobox  
CC zipper, LIM domain, AP2 and EREBS, zinc finger domains of type 2  
CC Cys2His2, CCAAT box elements and MYB.

SO Sequence 234 AA;

Query Match 61.5%; Score 8; DB 21; Length 234;  
Best Local Similarity 100.0%; Pred. No. 4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPA 11  
|||||||  
DB 35 PPPPLPPA 42

RESULT 12

AAB79117  
ID AAB79117 standard; Protein: 313 AA.

XX AC AAB79117;

XX DT 30-APR-2001 (first entry)

XX DE Corynebacterium glutamicum HA protein sequence SEQ ID NO:190.

XX KW Corynebacterium glutamicum; homeostasis; adaptation; HA protein;  
KW fine chemical production; organic acid; proteinogenic amino acid;  
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;  
KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;  
KW carboxylate; aromatic compound; vitamin; cofactor; polyketide; enzyme;  
KW diagnosis; Corynebacterium diptheriae; genetic engineering;  
KW Brevibacterium; environmental condition.

XX OS Corynebacterium glutamicum.

XX PN WO200100842-A2.

XX PD 04-JAN-2001.

XX PF 23-JUN-2000; 2000WO-IB00911.

XX PR 25-JUN-1999; 99US-0141031.

XX PR 08-JUL-1999; 99DE-1031636.

XX PR 09-JUL-1999; 99DE-1032125.

XX PR 09-JUL-1999; 99DE-1032126.

XX PR 09-JUL-1999; 99DE-1032127.

XX PR 09-JUL-1999; 99DE-1032128.

XX PR 09-JUL-1999; 99DE-1032129.

XX PR 09-JUL-1999; 99DE-1032226.

XX PR 14-JUL-1999; 99DE-1032920.

XX PR 14-JUL-1999; 99DE-1032922.

XX PR 14-JUL-1999; 99DE-1032924.

XX PR 14-JUL-1999; 99DE-1032928.

XX PR 14-JUL-1999; 99DE-1032930.

XX PR 14-JUL-1999; 99DE-1032933.

XX PR 14-JUL-1999; 99DE-1032935.

XX PR 14-JUL-1999; 99DE-1032973.

XX PR 14-JUL-1999; 99DE-1033002.

XX PR 14-JUL-1999; 99DE-1033003.

XX PR 14-JUL-1999; 99DE-1033005.

XX PR 14-JUL-1999; 99DE-1033006.

XX PR 31-AUG-1999; 99DE-1041378.

XX PR 31-AUG-1999; 99DE-1041379.

XX PR 31-AUG-1999; 99DE-1041380.

XX PR 31-AUG-1999; 99DE-1041391.

XX PR 03-SEP-1999; 99DE-1042088.

XX (BADI ) BASF AG.  
 XX Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;  
 PI WPI: 2001-061974/07.  
 DR N-PSDB: AAF71232.  
 XX  
 PT New isolated Corynebacterium glutamicum nucleic acid for production or  
 PT modulation of production of fine chemicals such as amino acids,  
 PT nucleosides, nucleotides, lipids, fatty acids, carbohydrates, vitamins  
 PT or enzymes -  
 PS Claim 20; Page 399-400; 712pp; English.  
 XX  
 CC AAF71138 to AAF71357 encode the Corynebacterium glutamicum homeostasis  
 CC and adaptation (HA) proteins given in AAB79023 to AAB79242. The  
 CC C. glutamicum HA genes (I) can be used in vectors for expression in host  
 CC cells and production of fine chemicals, such as, an organic acid,  
 CC proteinogenic or nonproteinogenic amino acid (preferred), purine or  
 CC pyrimidine base, nucleoside, nucleotide, lipid, saturated or unsaturated  
 CC fatty acid, diol, carbohydrate, aromatic compound, vitamin, cofactor,  
 CC polypeptide or enzyme. The amino acids produced can be lysine, glutamine,  
 CC glutamate, alanine, aspartate, glycine, serine, threonine, methionine,  
 CC cysteine, valine, leucine, isoleucine, arginine, proline, histidine,  
 CC tyrosine, phenylalanine, or tryptophan. The fine chemical production can  
 CC be modulated. The presence of (I) or HA proteins encoded by them are  
 CC used for diagnosing the presence or activity of Corynebacterium  
 CC diptheriae. (I) can be used to map the C. glutamicum genome or can be  
 CC used as markers for genetically engineered Corynebacterium or  
 CC Brevibacterium. The HA proteins encoded by the (I) are used to maintain  
 CC homeostasis in C. glutamicum or help the microorganism to adapt to  
 CC different environmental conditions.  
 SQ Sequence 313 AA;

Query Match 61.5%; Score 8; DB 22; Length 313;  
 Best Local Similarity 100.0%; Pred. No. 5;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 APPPLPP 10  
 |||||  
 Db 290 apppplpp 297

RESULT 13  
 AAB79242  
 ID AAB79242 standard; Protein; 313 AA.  
 XX  
 AC AAB79242;  
 XX  
 DT 30-APR-2001 (first entry)  
 XX  
 DE Corynebacterium glutamicum HA protein sequence SEQ ID NO:440.  
 XX  
 KW Corynebacterium glutamicum; homeostasis; adaptation; HA protein;  
 KW fine chemical production; organic acid; proteinogenic amino acid;  
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;  
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;  
 KW carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;  
 KW diagnosis; Corynebacterium diptheriae; genetic engineering;  
 KW Brevibacterium; environmental condition.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 PN WO200100842-A2.  
 XX  
 PD 04-JAN-2001.  
 XX  
 PF 23-JUN-2000; 2000MO-IB00911.  
 XX  
 PR 25-JUN-1999; 99US-0141031.

PR 08-JUL-1999; 99DE-1031636.  
 PR 09-JUL-1999; 99DE-1032125.  
 PR 09-JUL-1999; 99DE-1032126.  
 PR 09-JUL-1999; 99DE-1032127.  
 PR 09-JUL-1999; 99DE-1032128.  
 PR 09-JUL-1999; 99DE-1032129.  
 PR 09-JUL-1999; 99DE-1032226.  
 PR 14-JUL-1999; 99DE-1032920.  
 PR 14-JUL-1999; 99DE-1032922.  
 PR 14-JUL-1999; 99DE-1032924.  
 PR 14-JUL-1999; 99DE-1032928.  
 PR 14-JUL-1999; 99DE-1032930.  
 PR 14-JUL-1999; 99DE-1032933.  
 PR 14-JUL-1999; 99DE-1032935.  
 PR 14-JUL-1999; 99DE-1032973.  
 PR 14-JUL-1999; 99DE-1033002.  
 PR 14-JUL-1999; 99DE-1033003.  
 PR 14-JUL-1999; 99DE-1033005.  
 PR 14-JUL-1999; 99DE-1033006.  
 PR 31-AUG-1999; 99DE-1041378.  
 PR 31-AUG-1999; 99DE-1041379.  
 PR 31-AUG-1999; 99DE-1041390.  
 PR 31-AUG-1999; 99DE-1041391.  
 PR 03-SEP-1999; 99DE-1042088.  
 XX  
 PA (BADI ) BASF AG.  
 XX  
 PI Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;  
 DR WPI: 2001-061974/07.  
 XX N-PSDB: AAF71357.  
 PT  
 PT New isolated Corynebacterium glutamicum nucleic acid for production or  
 PT modulation of production of fine chemicals such as amino acids,  
 PT nucleosides, nucleotides, lipids, fatty acids, carbohydrates, vitamins  
 PT or enzymes -  
 PS Claim 20; Page 711-712; 712pp; English.  
 XX  
 CC AAF71138 to AAF71357 encode the Corynebacterium glutamicum homeostasis  
 CC and adaptation (HA) proteins given in AAB79023 to AAB79242. The  
 CC C. glutamicum HA genes (I) can be used in vectors for expression in host  
 CC cells and production of fine chemicals, such as, an organic acid,  
 CC proteinogenic or nonproteinogenic amino acid (preferred), purine or  
 CC pyrimidine base, nucleoside, nucleotide, lipid, saturated or unsaturated  
 CC fatty acid, diol, carbohydrate, aromatic compound, vitamin, cofactor,  
 CC polypeptide or enzyme. The amino acids produced can be lysine, glutamine,  
 CC glutamate, alanine, aspartate, glycine, serine, threonine, methionine,  
 CC cysteine, valine, leucine, isoleucine, arginine, proline, histidine,  
 CC tyrosine, phenylalanine, or tryptophan. The fine chemical production can  
 CC be modulated. The presence of (I) or HA proteins encoded by them are  
 CC used for diagnosing the presence or activity of Corynebacterium  
 CC diptheriae. (I) can be used to map the C. glutamicum genome or can be  
 CC used as markers for genetically engineered Corynebacterium or  
 CC Brevibacterium. The HA proteins encoded by the (I) are used to maintain  
 CC homeostasis in C. glutamicum or help the microorganism to adapt to  
 CC different environmental conditions.  
 SQ Sequence 313 AA;

Query Match 61.5%; Score 8; DB 22; Length 313;  
 Best Local Similarity 100.0%; Pred. No. 5;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 APPPLPP 10  
 |||||  
 Db 290 apppplpp 297

RESULT 14  
 AAG90541  
 ID AAG90541 standard; Protein; 438 AA.

XX AAG90541;  
 AC 26-SEP-2001 (first entry)  
 DT  
 XX C glutamicum protein fragment SEQ ID NO: 4295.  
 DE  
 XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
 KW organic acid synthesis.  
 KW  
 XX Corynebacterium glutamicum.  
 OS  
 XX EPI108790-A2.  
 PN  
 XX 20-JUN-2001.  
 PD  
 XX 18-DEC-2000; 2000EP-0127688.  
 PF  
 XX 16-DEC-1999; 99JP-0377484.  
 PR 07-APR-2000; 2000JP-0159162.  
 PR 03-AUG-2000; 2000JP-0280988.  
 XX  
 XX (KYOW ) KYOMA HAKKO KOGYO KK.  
 PA  
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
 PI  
 XX WPI: 2001-376931/40.  
 DR N-PSDB; AAH65760.  
 DR  
 XX Novel polynucleotides derived from Coryneform bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analysing  
 PT expression profile or pattern of a gene and identifying homologous gene  
 PT  
 XX Claim 17; SEQ ID NO: 4295; 246bp + Sequence Listing; English.  
 PS  
 XX The present invention provides a number of nucleotide and protein  
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of coryneform bacterium, measuring expression amount and  
 CC analysing the expression profile or expression pattern of a gene derived  
 CC from Coryneform bacterium, and identifying a homologue of a gene derived  
 CC from Coryneform bacterium. Coryneform bacteria are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a protein described  
 CC in the exemplification of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.  
 CC  
 XX Sequence 438 AA;  
 SQ

Query Match 61.5%; Score 8; DB 22; Length 438;  
 Best Local Similarity 100.0%; Pred. No. 6.6;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 3 APPPLPP 10  
 DB 314 appplpp 321

RESULT 15  
 AAB93086  
 ID AAB93086 standard; Protein; 586 AA.  
 XX  
 XX AAB93086;  
 AC  
 XX 26-JUN-2001 (first entry)  
 DT  
 XX Human protein sequence SEQ ID NO:11926.  
 DE

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 XX EPI1074617-A2.  
 PN  
 XX 07-FEB-2001.  
 PD  
 XX 28-JUL-2000; 2000EP-0116126.  
 PF  
 XX 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 XX (HELI-) HELIX RES INST.  
 PA  
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 PI  
 XX WPI: 2001-318749/34.  
 DR  
 XX  
 XX Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 PS  
 XX Claim 8; SEQ ID 11926; 2537bp + CD ROM; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification, where a primer set  
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03106 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 CC  
 XX Sequence 586 AA;  
 SQ

Query Match 61.5%; Score 8; DB 22; Length 586;  
 Best Local Similarity 100.0%; Pred. No. 8.4;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 4 PPPPLPPA 11  
 DB 18 pppplppa 25

Search completed: January 22, 2002, 16:31:44  
 Job time: 324 sec

---





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2002, 16:32:17 ; Search time 26.82 Seconds  
(without alignments)  
36.923 Million cell updates/sec

Title: US-09-485-529-103

Perfect score: 13

Sequence: 1 LNAPPPPLPPAPQ 13

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	61.5	443	2 T29820	hypothetical prote
2	8	61.5	564	2 T45866	hypothetical prote
3	8	61.5	907	2 E96636	hypothetical prote
4	8	61.5	1032	2 D83637	serine/threonine p
5	8	61.5	1520	2 T00273	hypothetical prote
6	7	53.8	138	2 C96734	hypothetical prote
7	7	53.8	145	2 T48552	glutaredoxin-like
8	7	53.8	219	2 T51382	achaete-scute homo
9	7	53.8	238	2 T32889	hypothetical prote
10	7	53.8	253	2 T17312	hypothetical prote
11	7	53.8	281	2 T38707	Fas ligand - human
12	7	53.8	285	2 T18689	hypothetical prote
13	7	53.8	289	2 T52354	hypothetical prote
14	7	53.8	341	2 D96500	probable transcrip
15	7	53.8	359	2 T13478	hypothetical prote
16	7	53.8	373	2 A70856	probable lppz prot
17	7	53.8	389	2 S27200	proline-rich prote
18	7	53.8	431	1 E70699	probable pknA prot
19	7	53.8	433	2 T07910	adenylyl cyclase-a
20	7	53.8	477	2 T38409	Wiskott-Aldrich sy
21	7	53.8	502	2 A55197	gag polyprotein -
22	7	53.8	538	1 F0MW1M	structural protein
23	7	53.8	555	2 T03049	probable low-affin
24	7	53.8	574	2 T05964	hypothetical prote
25	7	53.8	585	2 T00979	hypothetical prote
26	7	53.8	588	2 T45564	hypothetical prote
27	7	53.8	596	2 T03908	U25 protein - hum
28	7	53.8	656	1 Q0BET2	hypothetical prote
29	7	53.8	664	2 T01368	

30	7	53.8	678	2 T04832	probable serine/th
31	7	53.8	708	2 D96711	hypothetical prote
32	7	53.8	716	2 T25998	hypothetical prote
33	7	53.8	736	2 T51691	dishevelled homolo
34	7	53.8	820	2 T46412	ubiquitin--protein
35	7	53.8	933	1 ORHUP	progesterone recep
36	7	53.8	981	1 F0WVGM	gag-abl polyprotel
37	7	53.8	994	2 S19595	chloride channel p
38	7	53.8	998	2 T30930	hypothetical prote
39	7	53.8	1076	2 S50536	hypothetical prote
40	7	53.8	1110	2 T19673	hypothetical prote
41	7	53.8	1145	2 T18235	transcription acti
42	7	53.8	1206	2 S24407	formin isoform IV
43	7	53.8	1264	2 A36858	G2R protein - vari
44	7	53.8	1268	2 T31420	C-terminal domain
45	7	53.8	1344	2 T14316	r19-1 protein - mo

#### ALIGNMENTS

RESULT 1  
T29820  
hypothetical protein F25E2.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
R:Minx, P.; Le, T.  
C:Accession: T29820  
submitted to the EMBL Data Library, March 1996  
A:Description: The sequence of C. elegans cosmid F25E2.  
A:Reference number: Z20691  
A:Accession: T29820  
A:Status: preliminary; translated from GB/EMBL/DBBJ  
A:Molecule type: DNA  
A:Residues: 1-443 <MIN>  
A:Cross-references: EMBL:U50197; PIDN:AAA91257.1; CESP:F25E2.4  
A:Experimental source: strain Bristol N2  
C:Genetics:  
A:Gene: CESP:F25E2.4  
A:Introns: 72/3; 110/3; 143/2; 171/3; 330/3; 385/3

Query Match 61.5%; Score 8; DB 2; Length 443;  
Best Local Similarity 100.0%; Pred.No. 1.5;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 NAPPPLP 9  
DB 415 NAPPPLP 422

RESULT 2  
T45866  
hypothetical protein F3A4.210 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
R:Barques, M.; Collado, M.C.; Navarro, P.; Terol, J.; Perez-Alonso, M.; Mewes, H.W.;  
submitted to the Protein Sequence Database, December 1999  
A:Reference number: Z23007  
A:Accession: T45866  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-564 <BAR>  
A:Cross-references: EMBL:AL132978  
A:Experimental source: cultivar Columbia; BAC clone F3A4  
C:Genetics:  
A:Map position: 3  
A:Introns: 250/2; 424/1  
A:Note: F3A4.210

Query Match 61.5%; Score 8; DB 2; Length 564;

Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPA 11  
|||||  
Db 51 PPPPLPPA 58

## RESULT 3

E96636

hypothetical protein T7P1.21 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: E96636

R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Hultzer, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: E96636

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-907 &lt;STO&gt;

A:Cross-references: GB:AE005173; NID:96751696; PIDN:AAE27679.1; GSPDB:GN00141

C:Genetics:

A:Gene: T7P1.21

A:Map position: 1

Query Match 61.5%; Score 8; DB 2; Length 907;  
Best Local Similarity 100.0%; Pred. No. 2.8;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPA 11  
|||||  
Db 476 PPPPLPPA 483

## RESULT 4

D83637

serine/threonine protein kinase PpKa PA0074 [imported] - Pseudomonas aeruginosa (strain

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: D83637

R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B.

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim,

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: A82950; MUID:20437337

A:Accession: D83637

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1032 &lt;STO&gt;

A:Cross-references: GB:AE004446; GB:AE004091; NID:9945886; PIDN:AA03464.1; GSPDB:GN001

C:Genetics:

A:Experimental source: strain PA01

A:Gene: ppxA; PA0074

Query Match 61.5%; Score 8; DB 2; Length 1032;  
Best Local Similarity 100.0%; Pred. No. 3.2;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPA 11  
|||||  
Db 337 PPPPLPPA 344

RESULT 5  
T00273  
hypothetical protein KIAA0595 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 21-Jul-2000

C:Accession: T00273

R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara,

DNA Res. 5, 31-39, 1998

A:Title: Prediction of the coding sequences of unidentified human genes. IX. The comp

A:Reference number: Z14086; MUID:98290545

A:Accession: T00273

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1520 &lt;NMG&gt;

A:Cross-references: EMBL:AB011167; NID:93043713; PIDN:BA25521.1; PID:93043714

A:Experimental source: brain

C:Genetics:

A&gt;Note: KIAA0595

Query Match 61.5%; Score 8; DB 2; Length 1520;  
Best Local Similarity 100.0%; Pred. No. 4.5;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPA 11  
|||||  
Db 854 PPPPLPPA 861

## RESULT 6

C96734

hypothetical protein F15H11.18 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: C96734

R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,

ansen, N.F.; Hughes, B.; Hultzer, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marzia

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: C96734

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-138 &lt;STO&gt;

A:Cross-references: GB:AE005173; NID:95902409; PIDN:AAD5511.1; GSPDB:GN00141

C:Genetics:

A:Gene: F15H11.18

A:Map position: 1

Query Match 53.8%; Score 7; DB 2; Length 138;  
Best Local Similarity 100.0%; Pred. No. 4.2;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPP 10  
|||||  
Db 71 PPPPLPP 77

## RESULT 7

T48552

glutaredoxin-like protein - Arabidopsis thaliana

N:Alternate names: protein F14F18.100

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000

C:Accession: T48552  
R:Byan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft, submitted to the Protein Sequence Database, April 2000  
A:Reference number: 224490  
A:Accession: T48552  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-145 <BEV>  
A:Cross-references: EMBL:AL163812  
A:Experimental source: cultivar Columbia; BAC clone F14F18  
C:Genetics:  
A:Map position: 5  
A:Note: F14F18.100

Query Match 53.8%; Score 7; DB 2; Length 145;  
Best Local Similarity 100.0%; Pred. No. 4.4;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPP 10  
|||||  
DB 22 PPPPLPP 28

RESULT 8  
151382  
achaete-scute homolog - chicken  
C:Species: Gallus gallus (chicken)  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000  
C:Accession: 151382  
R:Jasoni, C.L.; Walker, M.B.; Morris, M.D.; Reh, T.A.  
Development 120, 769-783, 1994  
A:Title: A chicken achaete-scute homolog (CASH-1) is expressed in a temporally and spatially  
A:Reference number: 151382; MUID:95324365  
A:Accession: 151382  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-219 <JAS>  
A:Cross-references: EMBL:U01339; NID:g401726; PIDN:AAC59658.1; PID:g401727

Query Match 53.8%; Score 7; DB 2; Length 219;  
Best Local Similarity 100.0%; Pred. No. 6.3;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 APPPLPP 9  
|||||  
DB 42 APPPLPP 48

RESULT 9  
T32889  
hypothetical protein C34B2.9 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T32889  
R:Graves, T.; Sutterer, C.; Hawkins, M.; Wilson, R.  
submitted to the EMBL Data Library, January 1998  
A:Description: The sequence of C. elegans cosmid C34B2.  
A:Reference number: 221241  
A:Accession: T32889  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-238 <GRA>  
A:Cross-references: EMBL:AF043693; PIDN:AAB97541.1; GSPDB:GN00019; CESP:C34B2.9  
A:Experimental source: strain Bristol N2; clone C34B2  
C:Genetics:  
A:Gene: CESP:C34B2.9  
A:Map position: 1  
A:introns: 35/3; 91/2; 200/2

Query Match 53.8%; Score 7; DB 2; Length 238;

Best Local Similarity 100.0%; Pred. No. 6.8;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPP 10  
|||||  
DB 132 PPPPLPP 138

RESULT 10  
T17312  
hypothetical protein DKFZp434H018.1 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T17312  
R:Ottenwelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, September 1999  
A:Reference number: 218726  
A:Accession: T17312  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-253 <OT>  
A:Cross-references: EMBL:AL117579  
A:Experimental source: adult testis; clone DKFZp434H018  
C:Genetics:  
A:Note: DKFZp434H018.1

Query Match 53.8%; Score 7; DB 2; Length 253;  
Best Local Similarity 100.0%; Pred. No. 7.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPP 10  
|||||  
DB 35 PPPPLPP 41

RESULT 11  
138707  
Fas ligand - human  
C:Species: Homo sapiens (man)  
C:Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 21-Jul-2000  
C:Accession: 138707; JC2340; S57565; 138554  
R:Takahashi, T.; Tanaka, M.; Inazawa, J.; Abe, T.; Suda, T.; Nagata, S.  
Int. Immunol. 6, 1567-1574, 1994  
A:Title: Human Fas ligand: gene structure, chromosomal location and species specificity  
A:Reference number: 138707; MUID:95127560  
A:Accession: 138707  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-281 <RES>  
A:Cross-references: EMBL:U11821; NID:g595430; PIDN:AAC50124.1; PID:g595431  
R:Mita, E.; Hayashi, N.; Iio, S.; Takehara, T.; Hijioka, T.; Kasahara, A.; Fusamoto, Biochem. Biophys. Res. Commun. 204, 468-474, 1994  
A:Title: Role of Fas ligand in apoptosis induced by hepatitis C virus infection.  
A:Reference number: JC2340; MUID:95071350  
A:Accession: JC2340  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-281 <MIT>  
A:Cross-references: GB:D38122; DDBJ:D29820; NID:g601892; PIDN:BAA07320.1; PID:g136990  
R:Schatzlein, C.E.  
submitted to the EMBL Data Library, June 1995  
A:Reference number: S57565  
A:Accession: S57565  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-281 <SCH>  
A:Cross-references: EMBL:X89102; NID:g887455; PID:g887456  
R:Allderson, M.R.; Tough, T.W.; Davis-Smith, T.; Braddy, S.; Falk, B.; Schooley, K.A.; J. Exp. Med. 181, 71-77, 1995  
A:Title: Fas ligand mediates activation-induced cell death in human T lymphocytes.  
A:Reference number: 138554; MUID:95105731  
A:Accession: 138554  
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA  
A:Residues: 1-281 <RE2>  
A:Cross-references: EMBL:U08137; NID:g624627; PIDN:AAC50071.1; PID:g624628  
C:Genetics:  
A:Gene: FasL  
A:Introns: 151/1; 116/3  
C:Keywords: glycoprotein; transmembrane protein  
F:80-102/Domain: transmembrane #status predicted <TM>  
F:76,184,250/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 53.8%; Score 7; DB 2; Length 281;  
Best Local Similarity 100.0%; Pred. No. 8;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4 PPPPLPP 10  
|||||  
Db 50 PPPPLPP 56

RESULT 12  
118689  
hypothetical protein B0285.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T18689  
R:Stuison, J.  
submitted to the EMBL Data Library, June 1994

A:Reference number: Z19007

A:Accession: T18689

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-285 <MTI>

A:Cross-references: EMBL:Z34533; PIDN:CAA84294.1; GSPDB:GN00021; CESP:B0285.2

A:Experimental source: clone B0285

C:Genetics:

A:Gene: CESP:B0285.2

A:Map position: 3

A:Introns: 65/2; 131/2; 173/3; 211/3

Query Match 53.8%; Score 7; DB 2; Length 285;  
Best Local Similarity 100.0%; Pred. No. 8.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4 PPPPLPP 10  
|||||  
Db 207 PPPPLPP 213

RESULT 13  
152354  
hypothetical protein B11E6.30 [imported] - Neurospora crassa

C:Species: Neurospora crassa

C:Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 20-Oct-2000

C:Accession: T52354

R:Schulte, U.; Aign, V.; Hobeisel, J.; Brandt, P.; Farman, B.; Holland, R.; Nyakatura,  
submitted to the Protein Sequence Database, September 2000

A:Reference number: Z26053

A:Accession: T52354

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-289 <SCH>

A:Cross-references: EMBL:AL442043; GSPDB:GN00116; NCSP:B11E6.30

A:Experimental source: BAC clone B11E6; strain OR74A

C:Genetics:

A:Gene: NCSP:B11E6.30

A:Map position: 6

Query Match 53.8%; Score 7; DB 2; Length 289;

Best Local Similarity 100.0%; Pred. No. 8.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4 PPPPLPP 10  
|||||  
Db 43 PPPPLPP 49

RESULT 14

D96500

probable transcription factor [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: D96500

R:Thellogis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,  
ansen, N.F.; Hughes, B.; Hultzar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzla

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: AB6141; MUID:21016719

A:Accession: D96500

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-341 <STO>

A:Cross-references: GB:AE005173; NID:g7523681; PIDN:AAF63120.1; GSPDB:GN00141

C:Genetics:

A:Gene: F2J6.6

A:Map position: 1

Query Match 53.8%; Score 7; DB 2; Length 341;  
Best Local Similarity 100.0%; Pred. No. 9.5;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 LNAPPPP 7  
|||||  
Db 63 LNAPPPP 69

RESULT 15  
113478  
hypothetical protein 34F3.10 - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000

C:Accession: T13478

R:Valenti, P.; Salles, C.; Campbell, L.; Glover, D.

submitted to the EMBL Data Library, April 1999

A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.

A:Reference number: Z17685

A:Accession: T13478

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-359 <PHI>

A:Cross-references: EMBL:AL031583; NID:e1321005; PID:e1321018; PIDN:CA841346.1

C:Genetics:

A:Cross-references: FlyBase:FBgn0025623

A:Introns: 17/2; 50/3; 353/2

A:Note: EG:34F3.10

Query Match 53.8%; Score 7; DB 2; Length 359;

Best Local Similarity 100.0%; Pred. No. 9.9;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4 PPPPLPP 10  
|||||  
Db 170 PPPPLPP 176

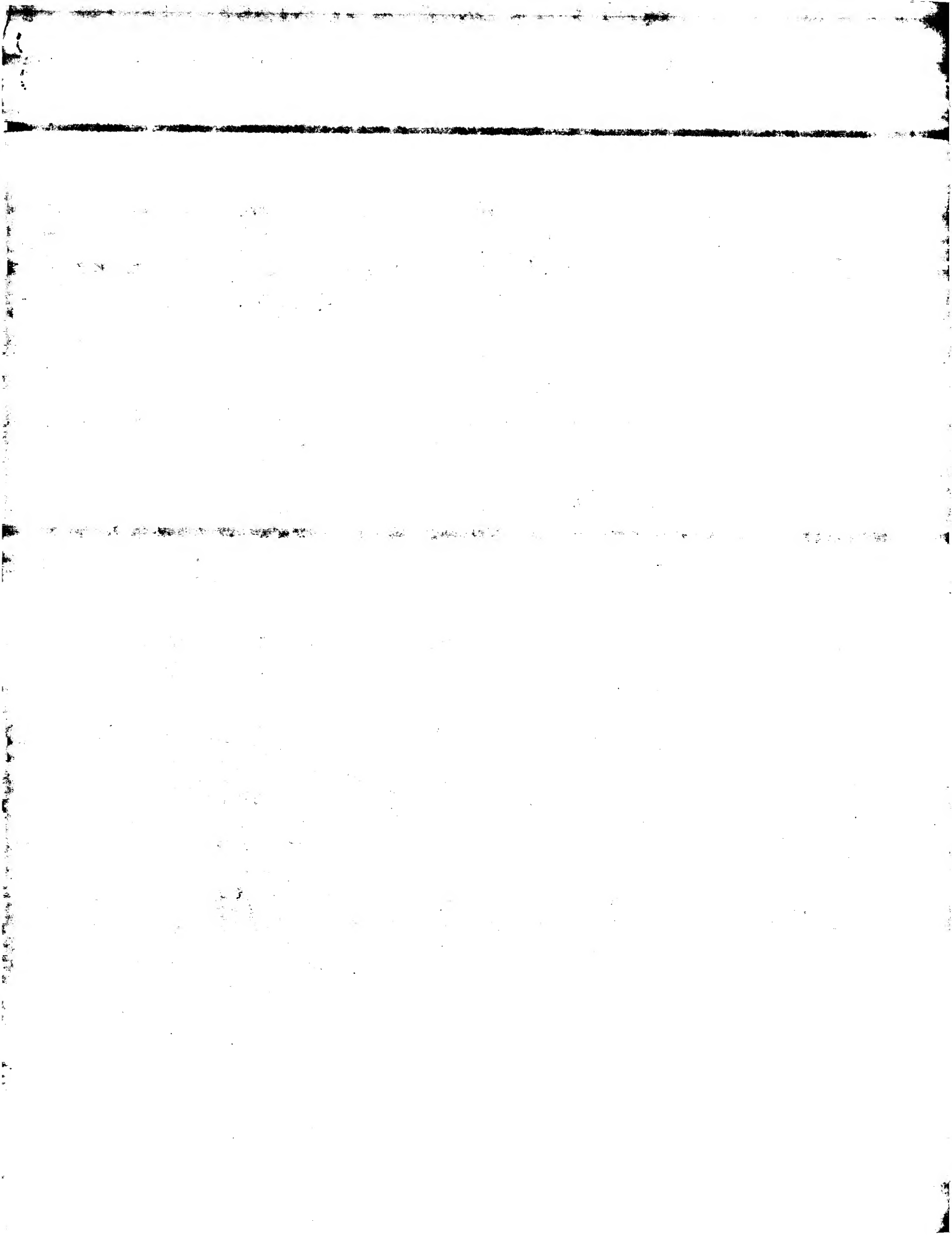
Thu Jan 24 09:56:01 2002

us-09-485-529-103.oli.rpr

Page 5

Search completed: January 22, 2002, 16:32:18  
Job time: 178 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2002, 16:37:08 ; Search time 17.09 Seconds  
(without alignments)  
27.890 Million cell updates/sec

Title: US-09-485-529-103  
Perfect score: 13  
Sequence: 1 LNAPPPLPPAPQ 13

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 100059 seqs, 36664827 residues

Word size : 0

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	69.2	282	1	ATF5_HUMAN
2	8	61.5	282	1	MTT_MOUSE
3	7	53.8	235	1	GAG_MLVAR
4	7	53.8	281	1	PASL_HUMAN
5	7	53.8	285	1	YK2_CAEEL
6	7	53.8	288	1	SMN_MOUSE
7	7	53.8	290	1	TRX2_MOUSE
8	7	53.8	314	1	Y009_HUMAN
9	7	53.8	314	1	PELB_COLGL
10	7	53.8	340	1	GBX2_CHICK
11	7	53.8	389	1	NDP_MOUSE
12	7	53.8	408	1	G3PT_HUMAN
13	7	53.8	412	1	ALF_PETHY
14	7	53.8	431	1	PKNA_MYCTU
15	7	53.8	453	1	GAA6_HUMAN
16	7	53.8	477	1	CAR2_HUMAN
17	7	53.8	502	1	WASP_HUMAN
18	7	53.8	538	1	GAG_MLVAR
19	7	53.8	559	1	WAS1_HUMAN
20	7	53.8	656	1	UL25_HCMVA
21	7	53.8	684	1	CDK9_CAEEL
22	7	53.8	724	1	CC11_MOUSE
23	7	53.8	726	1	CC11_MOUSE
24	7	53.8	727	1	CC11_MOUSE
25	7	53.8	736	1	DVL2_XENLA
26	7	53.8	736	1	DVL2_XENLA
27	7	53.8	757	1	CTRF_HUMAN
28	7	53.8	809	1	CNAD_HUMAN
29	7	53.8	884	1	ANDR_EDLFC
30	7	53.8	901	1	Y298_HUMAN
31	7	53.8	933	1	PRGR_HUMAN
32	7	53.8	971	1	Y029_HUMAN
33	7	53.8	994	1	CLC1_RAT

34	7	53.8	1076	1	YEM3_YEAST
35	7	53.8	1164	1	FHOS_HUMAN
36	7	53.8	1206	1	FM14_MOUSE
37	7	53.8	1468	1	FMN1_MOUSE
38	7	53.8	1733	1	VN1A_PPKA
39	7	53.8	1794	1	YAV1_SCHPO
40	7	53.8	1874	1	POIN_KYMT
41	7	53.8	2205	1	POLN_RUBVT
42	7	53.8	2440	1	NCR1_HUMAN
43	7	53.8	2453	1	NCR1_MOUSE
44	7	53.8	2715	1	TRX2_HUMAN
45	6	46.2	71	1	HSTA_YEREN

## ALIGNMENTS

RESULT 1  
ID ATF5\_HUMAN STANDARD: PRT; 282 AA.  
AC Q9Y2D1; Q9UNQ3;  
DT 20-AUG-2001 (Rel. 40, Created)  
DT 20-AUG-2001 (Rel. 40, last sequence update)  
DT 20-AUG-2001 (Rel. 40, last annotation update)  
DE CYCLIC-AMP-DEPENDENT TRANSCRIPTION FACTOR ATF-5 (ACTIVATING DE TRANSCRIPTION FACTOR 5) (TRANSCRIPTION FACTOR ATF5).  
GN ATF5 OR ATF5.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20558615; PubMed=11087824;  
RA White J.H., McIlhinney R.A.J., Wise A., Ciruela F., Chan W.Y., Emson P.C., Billington A., Marshall F.H.;  
RT "The GABA<sub>A</sub> receptor interacts directly with the related transcription factors CREB2 and ATF5.";  
RT Proc. Natl. Acad. Sci. U.S.A. 97:13967-13972(2000).  
RL [2]  
RN SEQUENCE FROM N.A.  
RA Kohroki J., Tanaka K.;  
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 161-282 FROM N.A.  
RX MEDLINE=99303793; PubMed=10373550;  
RA Pati D., Melstich M.L., Pion S.E.;  
RT "Human Cdc34 and Rad6 ubiquitin-conjugating enzymes target repressors of cyclic AMP-induced transcription for proteolysis.";  
RT Mol. Cell. Biol. 19:5001-5013(1999).  
RL  
CC -1- FUNCTION: THIS PROTEIN BINDS THE CAMP RESPONSE ELEMENT (CRE) (CONSENSUS: 5'GTGCGT(A/C)(A/G)-3'), A SEQUENCE PRESENT IN MANY VIRAL AND CELLULAR PROMOTERS.  
CC -1- SUBUNIT: BINDS DNA AS A DIMER (POTENTIAL). INTERACTS WITH GABAB RECEPTOR.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
CC -1- SIMILARITY: TO OTHER BZIP PROTEINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL: AF305687; AAG22558.1; -;  
DR EMBL: AB021653; BAA78477.2; -;  
DR EMBL: AF101588; AAD28570.1; -;  
DR InterPro: IPR002637; Hamip-like.  
DR InterPro: IPR002965; P-rich\_extensn.  
DR InterPro: IPR001871; bzip.  
DR Pfam: PF00170; bzip; 1.

DR	PRINTS; PR01217; PRICHEXTENSN.
DR	ProdDom; PD004952; HamP_1like; 1.
DR	SMART; SMO0338; BRZ; 1.
DR	PROSITE; PS00036; BZIP_BASIC; 1.
KW	Transcription regulation; DNA-binding; Activator; Nuclear protein;
KW	Multigene family.
FT	DOMAIN 123 139 POLY-PRO.
FT	DOMAIN 186 194 POLY-PRO.
FT	DNA_BIND 210 230 BASIC MOTIF.
FT	DOMAIN 236 250 LEUCINE-ZIPPER (REP.PROP).
FT	CONFLICT 161 163 LLA -> RHE (IN REF. 3)
SQ	SEQUENCE 282 AA; 30674 MW; DBB2F907CA0215A0 CRC64;
Query Match	
Best Local Similarity 69.2%; Score 9; DB 1; Length 282;	
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	4 PPPPLPPAP 12
db	131 PPPPLPPAP 139

RESULT	2	
ID	MNT_MOUSE	
AC	008789; P97349;	STANDARD; PRT: 591 AA.
DT	15-DEC-1998 (Rel. 37, Created)	
DT	15-DEC-1998 (Rel. 37, Last sequence update)	
DT	15-DEC-1998 (Rel. 37, Last annotation update)	
DE	MAX BINDING PROTEIN MNT (ROX PROTEIN) (MYC ANTAGONIST MNT).	
GN	MNT OR ROX.	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
OX	NCBI_TaxId=10090;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Embryo;	
RX	MEDLINE=97152466; PubMed=9000049;	
RA	Hurlin P.J., Queva C., Eisenman R.N.;	
RT	"Mut, a novel Max-interacting protein is coexpressed with Myc in	
RL	proliferating cells and mediates repression at Myc binding sites.;"	
RN	Genes Dev. 11:44-58(1997).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Embryo;	
RX	MEDLINE=97327566; PubMed=9184233;	
RA	Meroni G., Raymond A., Alcalay M., Borsani G., Tanigami A.,	
RA	Tonlorenzi R., Lo Negro C., Messali S., Zollo M., Ledbetter D.H.,	
RA	Brent R., Ballabio A., Carrozzo R.;	
RT	"Rox, a novel bHLHZip protein expressed in quiescent cells that	
RT	heterodimerizes with Max, binds a non-canonical E box and acts as a	
RT	transcriptional repressor.;"	
RL	EMBO J. 16:2892-2906(1997).	
CC	-1- FUNCTION: BINDS DNA AS A HETERODIMER WITH MAX AND REPRESENTS	
CC	TRANSCRIPTION. BINDS TO THE CANONICAL E BOX SEQUENCE 5'-CACCTG-3'	
CC	AND, WITH HIGHER AFFINITY, TO 5'-CACGCG-3'.	
CC	-1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER	
CC	BHLH PROTEIN. BINDS DNA AS AN HOMODIMER OR A HETERODIMER WITH MAX.	
CC	-1- SUBCELLULAR LOCATION: NUCLEAR.	
CC	-1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF	
CC	TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.	
CC	-----	
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> ).	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> .	
CC	-----	
CC	EMBL; U77356; AAB38687.1; -	

DR	EMLL; Y07609; CAA68878.1; -.
DR	MGD; MG1:109150; Mnt.
DR	InterPro; IPR003015; HLH_Myc.
DR	InterPro; IPR001092; HLH_dim.
DR	Pfam; PF00010; HLH; 1.
DR	SMART; SM00353; HLH; 1.
DR	PROSITE; PS00038; HELIX_LOOP_HELIX; FALSE_NEG.
KW	Transcription regulation; Repressor; Nuclear protein; DNA-binding.
FT	DNA_BIND 224 235
FT	DOMAIN 236 272
FT	DOMAIN 273 301
FT	CONFLICT 379 379
FT	CONFLICT 392 392
FT	CONFLICT 402 403
FT	CONFLICT 414 414
FT	CONFLICT 431 431
FT	CONFLICT 465 465
FT	CONFLICT 525 525
FT	CONFLICT 558 558
SQ	SEQUENCE 591 AA; 65311 MW; 3F4E61A10CFd142F CRC64;

[illegible]



```

FT CHAIN 2 131 CORE PROTEIN P15.
FT CHAIN 132 215 INNER COAT PROTEIN P12.
FT CHAIN 216 235 CORE SHELL PROTEIN P30.
FT LIPID 2 2 MYRISTATE.
SO SEQUENCE 235 AA; 25641 MW; 4D83F71D7E056C7D CRC64;

Query Match 53.8%; Score 7; DB 1; Length 235;
Best Local Similarity 100.0%; Pred. No. 3.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPP 10
DB 105 PPPPLPP 111

RESULT 4
FASL_HUMAN
ID FASL_HUMAN STANDARD; PRT; 281 AA.
AC P48023;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FAS ANTIGEN LIGAND (APOPTOSIS ANTIGEN LIGAND) (APTL).
GN TNFSF6 OR APTL1G OR FASL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RX MEDLINE=95105731; PubMed=7528780;
RA Alderson M.;
RT "Fas ligand mediates activation-induced cell death in human T
RT lymphocytes.";
RL J. Exp. Med. 181:71-77(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95127560; PubMed=7826947;
RA Takahashi T., Tanaka M., Inazawa J., Abe T., Suda T., Nagata S.;
RT "Human Fas ligand: gene structure, chromosomal location and species
RT specificity.";
RL Int. Immunol. 6:1567-1574(1994).
RN [3]
RP SEQUENCE FROM N.A.
RA Schaeublein C.E., Philippson P., Eibel H.;
RT Submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=95071350; PubMed=7980502;
RA Mita E., Hayashi N., Ito S., Takehana T., Hijioka T., Kasahara A.,
RA Fusamoto H., Kamada T.;
RT "Role of Fas ligand in apoptosis induced by hepatitis C virus
RT infection.";
RL Biochem. Biophys. Res. Commun. 204:468-474(1994).
RN [5]
RP SEQUENCE FROM N.A.
RA Wilkinson J.;
RT Submitted (MAY-1997) to the EMBL/Genbank/DBJ databases.
RN [6]
RP SEQUENCE OF 1-10 FROM N.A.
RC TISSUE=Blood;
RA Matsumura M., Nakanishi Y., Oiba Y.;
RT Submitted (APR-1998) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: CYTOKINE THAT BINDS TO FAS ANTIGEN, A RECEPTOR THAT
CC TRANSDUCES THE APOPTOTIC SIGNAL INTO CELLS. MAY BE INVOLVED IN
CC CYTOTOXIC T CELL MEDIATED APOPTOSIS AND IN T CELL DEVELOPMENT.
CC FAS ANTIGEN MEDIATED APOPTOSIS MAY HAVE A ROLE IN THE INDUCTION OF
CC PERIPHERAL TOLERANCE, IN THE ANTIGEN-STIMULATED SUICIDE OF MATURE
CC T CELLS, OR BOTH.
CC -1- SUBUNIT: HOMOTRIMER (PROBABLE).
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MAY BE RELEASED
CC INTO THE EXTRACELLULAR FLUID, PROBABLY BY CLEAVAGE FROM THE CELL

```

```

CC SURFACE.
CC -1- SIMILARITY: BELONGS TO THE TUMOR NECROSIS FACTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X89102; CAA61474.1; -
DR EMBL: U08137; AAC50071.1; -
DR EMBL: U11821; AAC50124.1; -
DR EMBL: D38122; BAA07320.1; -
DR EMBL: Z96050; CAB09424.1; -
DR EMBL: AB013303; BAA32542.1; -
DR HSSP: P01375; 2TUN.
DR MIM: 134638; -
DR InterPro: IPR003263; TNF_5.
DR InterPro: IPR000478; TNF_family.
DR Pfam: PF00229; TNF_1.
DR ProDom: PD008600; TNF_5; 1.
DR SMART: SM00207; TNF_1.
DR PROSITE: PS00251; TNF_1; 1.
DR PROSITE: PS50049; TNF_2; 1.
KW Cytokine; Transmembrane; Glycoprotein; Signal-anchor; Apoptosis.
FT DOMAIN 1 80 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 81 102 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT DOMAIN 103 281 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 4 70 PRO-RICH.
FT DISULFID 202 233 BY SIMILARITY.
FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE 281 AA; 31485 MW; A8A6B358246E9BB CRC64;

Query Match 53.8%; Score 7; DB 1; Length 281;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPP 10
DB 50 PPPPLPP 56

RESULT 5
YK62_CAEEL
ID YK62_CAEEL STANDARD; PRT; 285 AA.
AC P46552;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE HYPOTHEICAL 31.4 KDA PROTEIN B0285.2 IN CHROMOSOME III.
GN B0285.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Sulston J.;
RT Submitted (JUN-1994) to the EMBL/Genbank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch).

-----

CC EMBL; Z34533; CAA84294.1; -  
 DR Wormpep; B0285.2; CE00641.  
 KW Hypothetical protein.  
 FT DOMAIN 70 75 POLY-GLY.  
 FT DOMAIN 123 129 POLY-SER.  
 FT DOMAIN 163 170 POLY-SER.  
 FT DOMAIN 189 192 POLY-PRO.  
 FT DOMAIN 206 213 POLY-PRO.  
 FT DOMAIN 252 256 POLY-PRO.  
 SQ SEQUENCE 285 AA; 31396 MW; BFE68FF2038A3337 CRC64;

Query Match 53.8%; Score 7; DB 1; Length 285;  
 Best Local Similarity 100.0%; Pred. No. 4,7;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLP 10  
 |||||  
 DB 207 PPPPLP 213

RESULT 6  
 SMN\_MOUSE STANDARD; PRT; 288 AA.

AC P97801; O09092;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE SURVIVAL MOTOR NEURON PROTEIN.  
 GN SMN1 OR SMN.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9724505; PubMed=9070939;  
 RA Viollet L., Bertrand S., Brunialti A.L.B., Lefebvre S., Burtet P.,  
 RA Clermont O., Cruaud C., Guenet J.-L., Munnich A., Melki J.,  
 RT "cDNA isolation, expression, and chromosomal localization of the  
 RT mouse survival motor neuron gene (Smn).";  
 RL Genomics 40:185-188(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97264340; PubMed=9110173;  
 RA Diconato C.J., Chen X.N., Noya D., Korenberg J.R., Nadeau J.H.,  
 RA Simard L.R.;  
 RT "Cloning, characterization, and copy number of the murine survival  
 RT motor neuron gene: homolog of the spinal muscular atrophy-determining  
 RT gene.";  
 RL Genome Res. 7:339-352(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN-BALB/C; TISSUE-Brain;  
 RX MEDLINE=97420785; PubMed=9275227;  
 RA Schrank B., Goetz R., Gunnarsen J.M., Ure J.M., Toyka K.V.,  
 RA Smith A.G., Sendtner M.;  
 RT "Inactivation of the survival motor neuron gene, a candidate gene for  
 RT human spinal muscular atrophy, leads to massive cell death in early  
 RT mouse embryos.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:9920-9925(1997).  
 CC -1- FUNCTION: THE SMN COMPLEX PLAYS AN ESSENTIAL ROLE IN SPLICESOMAL  
 CC SNRP ASSEMBLY IN THE CYTOPLASM, AND IS REQUIRED FOR PRE-MRNA  
 CC SPLICING IN THE NUCLEUS (BY SIMILARITY).  
 CC -1- SUBUNIT: FORMS A STABLE HETEROMERIC COMPLEX WITH SURVIVAL OF MOTOR  
 CC NEURON PROTEIN INTERACTING PROTEIN 1 (SIP1/GEMIN2), GEMIN3 AND  
 CC GEMIN4. THIS COMPLEX IS THOUGHT TO PLAY AN ESSENTIAL ROLE IN  
 CC BIOGENESIS OF THE SMALL NUCLEAR RIBONUCLEOPROTEIN PARTICLES  
 CC (SNRNPs). INTERACTS WITH FIBRILLARIN AND WITH SEVERAL SPLICESOMAL  
 CC SNRP CORE SM PROTEINS, INCLUDING B/B', SM D AND SM E PROTEINS (BY  
 CC SIMILARITY).

-1- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR. LOCALIZED IN  
 CC SUBNUCLEAR STRUCTURES NEXT TO COILED BODIES, CALLED GEMs, WHICH  
 CC ARE HIGHLY ENRICHED IN SPLICESOMAL SNRNPs (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE SMN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.lsb-sib.ch/announce/>  
 CC or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).  
 CC -----

CC EMBL; U63294; AAC53057.1; -  
 DR EMBL; U77714; AAC53144.1; -  
 DR EMBL; Y12835; CAA73356.1; -  
 DR MGD; MGI:109257; Smn.  
 DR InterPro; IPR002999; Tudor.  
 DR SMART; SM00333; Tudor, 1.  
 KW mRNA processing; RNA-binding; Nuclear protein.  
 FT DOMAIN 190 196 POLY-PRO.  
 FT DOMAIN 212 222 POLY-PRO.  
 FT DOMAIN 239 243 POLY-PRO.  
 SQ SEQUENCE 288 AA; 31254 MW; 757B3074649F7458 CRC64;

Query Match 53.8%; Score 7; DB 1; Length 288;  
 Best Local Similarity 100.0%; Pred. No. 4,7;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLP 10  
 |||||  
 DB 214 PPPPLP 220

RESULT 7  
 TRX2\_MOUSE STANDARD; PRT; 290 AA.

AC O08550;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE TRITHORAX HOMOLOG 2 (TW DOMAIN BINDING PROTEIN 7) (FRAGMENT).  
 GN TRX2 OR WBP7.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97315177; PubMed=9171351;  
 RA Bedford M.T., Chan D.C., Leder P.;  
 RT "RBP W domains and the Abi SH3 domain bind to a specific class of  
 RT proline-rich ligands.";  
 RL EMBO J. 16:2376-2383(1997).  
 CC -1- FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE TRITHORAX FAMILY OF TRANSCRIPTION  
 CC FACTORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.lsb-sib.ch/announce/>  
 CC or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).  
 CC -----

CC EMBL; U92455; AAC53192.1; -  
 DR MGD; MGI:109565; Wbp7  
 DR InterPro; IPR000637; AT\_hook.  
 DR Pfam; PF02178; AT\_hook; 1.  
 DR SMART; SM00384; AT\_hook; 1.

KM DNA-binding: Nuclear Protein; Transcription regulation.  
 FT NON\_TER 1 1  
 FT DNA\_BIND <1 7  
 FT DOMAIN 4 35 A.T HOOK (BY SIMILARITY).  
 FT DOMAIN 45 77 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 253 263 POLY-PRO.  
 FT NON\_TER 290 290 POLY-PRO.  
 SO SEQUENCE 290 AA; 31202 MW; C49B5565F8CD2DF CRC64;

Query Match  
 Best Local Similarity 100.0%; Pred. No. 4.8;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLP 10  
 |||||  
 Db 50 PPPPLP 56

RESULT 8  
 Y009\_HUMAN STANDARD; PRT; 314 AA.  
 AC Q15390;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE HYPOTHETICAL PROTEIN KIAA0009.  
 GN KIAA0009.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Bone marrow;  
 RX MEDLINE=96051387; PubMed=7584026;  
 RA Nomura N., Miyajima N., Sazuka T., Tanaka A., Kawarabayashi Y.,  
 RA Sato S., Nagase T., Seki N., Ishikawa K.-I., Tabata S.;  
 RT "Prediction of the coding sequences of unidentified human genes. I.  
 RT The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by  
 RT analysis of randomly sampled cDNA clones from human immature myeloid  
 RT cell line KG-1."  
 RL DNA Res. 1:27-35(1994).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL: D13634; BAA02798.1; -.  
 KM HYPOTHETICAL protein  
 FT DOMAIN 165 178 POLY-PRO.  
 SO SEQUENCE 314 AA; 34756 MW; E79E4FE31061953 CRC64;

Query Match  
 Best Local Similarity 100.0%; Pred. No. 5.1;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLP 10  
 |||||  
 Db 171 PPPPLP 177

RESULT 9  
 PBLB\_COLGL STANDARD; PRT; 331 AA.  
 AC 059939;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE PEPTATE LYASE B PRECURSOR (EC 4.2.2.2).  
 GN PEB.  
 OS Colletotrichum gloeosporioides (Anthracnose fungus) (Glomerella  
 OS cingulata).  
 OC Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Sordariomycetes incertae sedis; Phyllachorales; Phyllachoraceae;  
 OC Glomerella.  
 OX NCBI\_TaxID=5457;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CG-14;  
 RA Mettad C., Keen N.T.;  
 RU Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: ACTS AS A VIRULENCE FACTOR ACTIVE IN PLANT TISSUE  
 CC MACERATION.  
 CC -1- CATALYTIC ACTIVITY: ELIMINATIVE CLEAVAGE OF PEPTATE TO GIVE  
 CC OLIGOSACCHARIDES WITH 4-DEOXY-ALPHA-D-MANN-4-ENURONOSYL GROUPS  
 CC AT THEIR NON-REDUCING ENDS.  
 CC -1- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).  
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL: AF052632; AAD09857.1; -.  
 DR InterPro: IPR002022; Amb\_allergen.  
 DR Pfam: PF00544; pec\_lyase; 1.  
 DR KW Lyase; Signal.  
 FT SIGNAL 1 2 POTENTIAL.  
 FT CHAIN 99 99 PEPTATE LYASE B.  
 FT CARBOHYD 99 99 N-LINKED (GLCNAC... ) (POTENTIAL).  
 SO SEQUENCE 331 AA; 35108 MW; 475A76F25D4689 CRC64;

Query Match  
 Best Local Similarity 100.0%; Pred. No. 5.4;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 APPPLP 9  
 |||||  
 Db 18 APPPLP 24

RESULT 10  
 GBX2\_CHICK STANDARD; PRT; 340 AA.  
 ID GBX2\_CHICK  
 AC 042230;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE HOMEOBOX PROTEIN GBX-2 (GASTRULATION AND BRAIN-SPECIFIC HOMEOBOX  
 DE PROTEIN 2).  
 GN GBX2.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98004222; PubMed=9346236;  
 RA Kowenz-Leutz E., Herr P., Niss K., Leutz A.;  
 RT "The homeobox gene GBX2, a target of the myb oncogene, mediates  
 RT autocrine growth and monocytic differentiation.";  
 RL Cell 91:185-195(1997).  
 CC -1- FUNCTION: MAY ACT AS A TRANSCRIPTION FACTOR. FOR CELL PLURIPOTENCY  
 CC AND DIFFERENTIATION IN THE EMBRYO.



SEQ SEQUENCE 408 AA; 44501 MW; 301F71C768CD95D8 CRC64;

Query Match 53.8%; Score 7; DB 1; Length 408;  
Best Local Similarity 100.0%; Pred. No. 6.5;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPP 10  
| | | | |  
Db 54 PPPPLPP 60

## RESULT 13

ALF\_PETHY STANDARD; PRT; 412 AA.  
AC 022621;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DE 30-MAY-2000 (Rel. 39, Last annotation update)  
DE ALF PROTEIN (ABERRANT LEAF AND FLOWER PROTEIN).  
GN ALF.  
OS Petunia hybrida (Petunia).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids I; Solanales; Solanaceae; Petunia.  
OX NCBI\_TaxID=4102;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. W138;  
RC MEDLINE=98167899; PubMed=9435293;  
RA Souer E., van der Krol A., Kloos D., Spelt C., Bliek M., Mol J.,  
RA Koes R.;  
RT "Genetic control of branching pattern and floral identity during  
RT Petunia inflorescence development.";  
RL Development 125:733-742(1998).  
CC -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR REQUIRED FOR THE  
CC SPECIFICATION OF FLORAL MERISTEM IDENTITY.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).  
CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE FLORAL MERISTEM AND ALSO IN  
CC THE VEGETATIVE MERISTEM.  
CC -1- SIMILARITY: BELONGS TO THE FLO / LEY FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: AF030171; AAC49912.1; -;  
DR Mendeil; 25747; Pctc;1429;25747.  
DR InterPro: IPR002910; FLO\_LEY.  
DR Pfam: PF01698; FLO\_LEY; 1.  
KW Transcription regulation; Activator; DNA-binding;  
KW Nuclear protein; Developmental protein.  
FT DOMAIN 31 44 POLY-PRO.  
FT DOMAIN 129 132 POLY-GLU.  
FT DOMAIN 169 174 POLY-GLY.  
FT DOMAIN 228 231 POLY-GLY.  
FT DOMAIN 393 397 POLY-ALA.  
SQ SEQUENCE 412 AA; 46268 MW; F1C02F8E95AB33 CRC64;

Query Match 53.8%; Score 7; DB 1; Length 412;  
Best Local Similarity 100.0%; Pred. No. 6.5;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPP 10  
| | | | |  
Db 35 PPPPLPP 41

RESULT 14  
ID PKNA\_MYCTU STANDARD; PRT; 431 AA.  
AC P71565;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DE 20-AUG-2001 (Rel. 40, Last annotation update)  
DE PROBABLE SERINE/THREONINE-PROTEIN KINASE PKNA (EC 2.7.1.-).  
GN PKNA OR RV0015C OR MT0018 OR MT0104.15C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RC MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence.";  
RL Nature 393:537-544(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RC Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Esmailova M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains.";  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: Z80233; CAB02435.1; -;  
DR EMBL: AE006916; AAK44240.1; -;  
DR TIGR: MT0018; -;  
DR Tuberculist; RV0015C; -;  
DR InterPro: IPR000719; Euk\_pkinase.  
DR InterPro: IPR002290; Ser\_thr\_kin\_actsite.  
DR Pfam: PF00069; pkinase; 1.  
DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; FALSE-NEG.  
DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
KW Transferase; Serine/threonine-protein kinase; ATP-binding;  
KW Complete proteome.  
FT DOMAIN 13 253 PROTEIN KINASE.  
FT NP\_BIND 19 27 ATP (BY SIMILARITY).  
FT BINDING 42 42 ATP (BY SIMILARITY).  
FT ACT\_SITE 141 141 ATP (BY SIMILARITY).  
SQ SEQUENCE 431 AA; 45597 MW; 582D183747F3C111 CRC64;

Query Match 53.8%; Score 7; DB 1; Length 431;  
Best Local Similarity 100.0%; Pred. No. 6.8;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPP 10

Db 233 PPPPLPP 239

RESULT: 15

GAA6\_HUMAN STANDARD; PRT; 453 AA.

AC 016445:

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE GAMMA-AMINOBUTYRIC-ACID RECEPTOR ALPHA-6 SUBUNIT PRECURSOR (GABA(A) RECEPTOR).

DE GABRA6.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

OX NCBI\_Taxid=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Cerebellum;

RX MEDLINE=96226062; PubMed=8632757;

RA Hadingham K.L., Garrett E.M., Mafford K.A., Bain C., Heavens R.P.,

Stinachsinghi D.J., Whiting P.J.;

RT Cloning of cDNAs encoding the human gamma-aminobutyric acid type A

receptor alpha 6 subunit and characterization of the pharmacology of

alpha 6-containing receptors.;

Mol. Pharmacol. 49:253-259(1996).

CC -! FUNCTION: GABA, THE MAJOR INHIBITORY NEUROTRANSMITTER IN THE

VERTEBRATE BRAIN, MEDIATES NEURONAL INHIBITION BY BINDING TO THE

GABA/BENZODIAZEPINE RECEPTOR AND OPENING AN INTEGRAL CHLORIDE

CHANNEL.

CC -! SUBUNIT: GENERALLY PENTAMERIC. THERE ARE FIVE TYPES OF GABA(A)

RECEPTOR CHAINS: ALPHA, BETA, GAMMA, DELTA, AND RHO.

CC -! SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

CC -! SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL: S81944; AAB36480.1; -.

CC MIM: 137143; -.

DR InterPro: IPR001175; Neur\_channel.

DR Pfam: PF00065; neur\_chan; 2.

DR PRINTS: PRO0252; NRIONCHANNEL.

DR PRINTS: PRO0253; GABARECEPTR.

DR PRINTS: PRO1079; GABARALPHA.

DR PROSITE: PS00236; NEUROTR\_ION\_CHANNEL; 1.

KW Postsynaptic membrane; Ionic channel; Glycoprotein; signal;

KW Transmembrane; Multigene family.

FT SIGNAL 1 19 POTENTIAL.

FT CHAIN 20 453 GAMMA-AMINOBUTYRIC-ACID RECEPTOR ALPHA-6

FT SUBUNIT. SUBUNIT.

FT DOMAIN 20 242 EXTRACELLULAR (PROBABLE).

FT TRANSMEM 243 264 PROBABLE.

FT TRANSMEM 269 290 PROBABLE.

FT TRANSMEM 301 324 PROBABLE.

FT DOMAIN 325 419 CYTOPLASMIC (PROBABLE).

FT TRANSMEM 420 441 PROBABLE.

FT CARBOHYD 31 31 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT DISULFD 156 170 BY SIMILARITY.

SO SEQUENCE 453 AA: 51034 MW: 780F85845BCB107F CRC64;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 5 PPPPLPPA 11  
| | | | |  
DB 404 PPPPLPPA 410

Search completed: January 22, 2002, 16:37:08  
Job time: 318 sec

Query Match 53.8%; Score 7; DB 1; Length 453;  
Best Local Similarity 100.0%; Pred. No. 7.1;







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2002, 16:36:44 ; Search time 40.87 Seconds  
(without alignments)  
46.527 Million cell updates/sec

Title: US-09-485-529-103  
Perfect score: 13  
Sequence: 1 LNAPPLPPAPQ 13

Scoring table:  
OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 473505 seqs, 146272329 residues

Word size : 0

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL\_17:\*  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp.protent:\*  
12: sp.virus:\*  
13: sp.vertebrate:\*  
14: sp.unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13	100.0	623	10	Q9ST59
2	9	69.2	282	4	Q9BSA1
3	9	69.2	407	2	Q9AAS9
4	8	61.5	66	10	Q9FT15
5	8	61.5	108	5	Q9VNZ6
6	8	61.5	179	11	Q9VMA0
7	8	61.5	195	10	Q9LW68
8	8	61.5	309	5	Q9YV01
9	8	61.5	443	5	Q19782
10	8	61.5	451	11	Q62775
11	8	61.5	453	4	Q95763
12	8	61.5	485	11	Q920G8
13	8	61.5	507	10	Q9FU62
14	8	61.5	564	10	Q9SN05
15	8	61.5	569	5	Q9W1X3
16	8	61.5	586	4	Q9NV17
17	8	61.5	648	4	Q9BUK4
18	8	61.5	770	5	Q44014
19	8	61.5	804	4	Q9UPS6

20	8	61.5	816	5	Q97214	Q97214 leishmania
21	8	61.5	862	5	Q44708	Q44708 drosophila
22	8	61.5	907	10	Q9C946	Q9C946 arabidopsis
23	8	61.5	956	5	Q9G021	Q9G021 drosophila
24	8	61.5	1032	2	Q921B2	Q921B2 pseudomonas
25	8	61.5	1032	2	Q91758	Q91758 pseudomonas
26	8	61.5	1056	5	Q9V126	Q9V126 drosophila
27	8	61.5	1520	4	Q9Y4E0	Q9Y4E0 homo sapien
28	8	61.5	1544	4	Q9NRJ4	Q9NRJ4 homo sapien
29	8	61.5	1547	11	Q9TIL5	Q9TIL5 mus musculu
30	8	61.5	1614	5	Q9BN25	Q9BN25 leishmania
31	8	61.5	1664	4	Q9BZE5	Q9BZE5 homo sapien
32	8	61.5	4833	11	Q9QYX6	Q9QYX6 mus musculu
33	8	61.5	5038	11	Q9QYX7	Q9QYX7 mus musculu
34	7	53.8	39	4	Q9BR93	Q9BR93 homo sapien
35	7	53.8	61	11	Q08349	Q08349 mus musculu
36	7	53.8	127	4	Q9B2P9	Q9B2P9 homo sapien
37	7	53.8	138	10	Q9SKR0	Q9SKR0 arabidopsis
38	7	53.8	145	10	Q9LY10	Q9LY10 mus musculu
39	7	53.8	149	11	Q93327	Q93327 mus musculu
40	7	53.8	171	4	Q9NY88	Q9NY88 homo sapien
41	7	53.8	179	4	Q9BTW1	Q9BTW1 homo sapien
42	7	53.8	180	5	Q901B2	Q901B2 leishmania
43	7	53.8	184	10	Q9XHM5	Q9XHM5 oryza sativ
44	7	53.8	210	4	Q9NXE9	Q9NXE9 homo sapien
45	7	53.8	213	4	Q9NVV9	Q9NVV9 homo sapien

## ALIGNMENTS

RESULT 1					
Q9ST59		PRELIMINARY;	PRT;	623	AA.
AC Q9ST59;					
DT 01-MAY-2000 (TREMBLrel. 13, Created)					
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)					
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)					
DE GIBBERELLIN RESPONSE MODULATOR.					
GN RHT-DIA.					
OS Triticum aestivum (wheat).					
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;					
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;					
OC Triticaceae; Triticum.					
OX NOBI_TaxID=4565;					
RN [1]					
RP SEQUENCE FROM N.A.					
RA MEDLINE-99347734; PubMed-10421366;					
RA Best J., Richards D.E., Hartley N.M., Murphy G.P., Devos K.M.,					
RA Flintham J.E., Beales J., Fish L.J., Worland A.J., Pelica F.,					
RA Sudhakar D., Christou P., Snape J.W., Gale M.D., Harberd N.P.;					
RT "Green revolution" genes encode mutant gibberellin response					
RT modulators.";					
RL Nature 400:256-261(1999).					
DR EMBL: AJ242531; CAB51555.1; .					
SQ SEQUENCE 623 AA; 65337 MW; F2BNC34996D0A84F CRC64;					
Query Match	100.0%;	Score 13;	DB 10;	Length 623;	
Best local Similarity	100.0%;	Pred. No. 1.9e-05;			
Matches 13;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
OY 1 LNAPPLPPAPQ 13					
DB 109 LNAPPLPPAPQ 121					
RESULT 2					
Q9BSA1		PRELIMINARY;	PRT;	282	AA.
ID Q9BSA1;					
AC Q9BSA1;					
DT 01-JUN-2001 (TREMBLrel. 17, Created)					
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)					

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE ACTIVATING TRANSCRIPTION FACTOR 5.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euteria; Primates; Catarrhini; Homidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=CHORIOCARCINOMA;  
RA Strassberg R.;  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC005174; AA05174.1; -  
SQ SEQUENCE 282 AA; 30690 MW; B6D4CB6F631655B1 CRC64;

Query Match 69.2%; Score 9; DB 4; Length 282;  
Best Local Similarity 100.0%; Pred. No. 0.082;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPAP 12  
|||||  
DB 131 PPPPLPPAP 139

RESULT 3  
O9AA59 PRELIMINARY; PRT; 407 AA.  
AC O9AA59:  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE OMPA FAMILY PROTEIN.  
GN CC0747.  
OS Caulobacter crescentus.  
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
OC Caulobacter.  
OX NCBI\_TaxID=69394;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21173698; Pubmed=11259647;  
RA Nieman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,  
Kutler J.F., Smit J., Craven M.B., Khouli H., Shetty J., Berry K.,  
Rajkumar T., Tran K., Wolf A., Yamahara J., Ermolaeva M., White O.,  
Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
RT "Complete genome sequence of Caulobacter crescentus."  
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
DR EMBL; AE005750; AAK22732.1; -  
DR TIGR; CC0747; -  
KM Complete proteome.  
SQ SEQUENCE 407 AA; 42781 MW; 54610D5FA64CC3BF CRC64;

Query Match 69.2%; Score 9; DB 2; Length 407;  
Best Local Similarity 100.0%; Pred. No. 0.11;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPAP 12  
|||||  
DB 260 PPPPLPPAP 268

RESULT 4  
O9FT15 PRELIMINARY; PRT; 66 AA.  
AC O9FT15:  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE P0410E01.15 PROTEIN.  
GN P0410E01.15.

OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrharioideae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. NIPPONBARE;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
clone:P0410E01."  
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AP002866; BAB17094.1; -  
SQ SEQUENCE 66 AA; 7427 MW; 8D5E3D01574E867C CRC64;

Query Match 61.5%; Score 8; DB 10; Length 66;  
Best Local Similarity 100.0%; Pred. No. 0.24;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPA 11  
|||||  
DB 4 PPPPLPPA 11

RESULT 5  
O9VNZ6 PRELIMINARY; PRT; 108 AA.  
AC O9VNZ6:  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
DE CG14571 PROTEIN.  
GN CG14571.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; Pubmed=10731132;  
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
Amaralidis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
Georgie R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
Dodson K., Dong L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
Hostlin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,  
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
Liu X., Matel B., McIntosh T.C., McLeod M.P., McPherson D.,  
Merkulov G., Milshina N.V., Mobarry C., Morris J., Mosheiff A.,  
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauley J.M.,  
Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Massaman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Morley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
DR EMBL: AE003595; AAF51766.1; -  
DR FlyBase: FBgn0037118; CG14571.  
SQ SEQUENCE 108 AA; 11926 MW; B67F87ACB2B3A45 CRC64;

Query Match 61.5%; Score 8; DB 5; Length 108;  
Best Local Similarity 100.0%; Pred. No. 0.36;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 PPPPLPP 10  
Db 20 APPPLPP 27

RESULT 6  
O99MA0 PRELIMINARY; PRT; 179 AA.  
AC O99MA0:  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE TRANSCRIPTION FACTOR MRG2.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=HEART;  
RA Zhang D.Z., Chou Y.-T., Yang Y.-C.;  
RT "Structural and functional conservation of MRG family in system  
evolution.";  
RL EMBL: AF361477; AAK30622.1; -  
DR SUBMIT: AF361477; AAK30622.1; -  
SQ SEQUENCE 179 AA; 18151 MW; 7B1A819DE7865284 CRC64;

Query Match 61.5%; Score 8; DB 11; Length 179;  
Best Local Similarity 100.0%; Pred. No. 0.54;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPP 11  
Db 116 PPPPLPP 123

RESULT 7  
O91M68 PRELIMINARY; PRT; 195 AA.  
AC O91M68:  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE GB|AC63835.1.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.,  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
[2]

RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RX MEDLINE=20277480; PubMed=10819329;  
RA Nakamura Y.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence  
features of the regions of 4,504,864 bp covered by sixty pl and YAC  
clones.";  
RL DNA Res. 7:131-135(2000).  
DR EMBL: AB015474; BAB02272.1; -  
DR EMBL: AP00733; BAB02272.1; JOINED.  
SQ SEQUENCE 195 AA; 21434 MW; A6BA85958BEFCE23 CRC64;

Query Match 61.5%; Score 8; DB 10; Length 195;  
Best Local Similarity 100.0%; Pred. No. 0.58;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPP 11  
Db 179 PPPPLPP 186

RESULT 8  
O9YV01 PRELIMINARY; PRT; 309 AA.  
AC O9YV01:  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
DE CG13055 PROTEIN.  
GN CG13055.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Abmayyan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegun C.,  
RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy J., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puti V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,  
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Massaman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Morley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaverl J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003528; AAF9521.1; -;  
 DR FlyBase: FBgn0036583; CC13055.  
 SQ SEQUENCE 309 AA; 33224 MW; 9DAEB67784852A93 CRC64;

Query Match 61.5%; Score 8; DB 5; Length 309;  
 Best Local Similarity 100.0%; Pred. No. 0.84;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPA 11  
 |||||  
 DB 211 PPPPLPPA 218

RESULT 9  
 019782 PRELIMINARY; PRT; 443 AA.  
 AC 019782;

DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE SIMILAR TO CYTOPLASMIC INTERMEDIATE FILAMENT, NCBI GI: 1213546.  
 GN F25B2.4.

OS *Caenorhabditis elegans*.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodetidae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RX MEDLINE=94150718; PubMed=7906398;

RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Cratton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
 RA Gardner A., Green P., Hawkins J., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Kershaw J., Kirsten J., Lalster N., Latreille P.,  
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifkin L., Koopra A., Saunders D., Showkseen R.,  
 RA Smaildon N., Smith A., Sonnenhammer E., Staden R., Sulston J.,  
 RA Thierly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Woldman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C.*  
 RT *elegans*.";  
 RL Nature 368:32-38(1994).  
 RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA Mink P., Le T.;

RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL: U50197; AAA91257.1; -;  
 DR InterPro: IPR001664; IF.

DR Pfam: PF00038; filament; 2.

SQ SEQUENCE 443 AA; 51581 MW; 2B6A54DDCB17730B CRC64;

Query Match 61.5%; Score 8; DB 5; Length 443;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 NAPPPLP 9  
 |||||  
 DB 415 NAPPPLP 422

RESULT 10  
 062775 PRELIMINARY; PRT; 451 AA.  
 AC 062775;

DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE SH3 DOMAIN BINDING PROTEIN.  
 GN CR16.

OS *Rattus norvegicus* (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.

OX NCBI\_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-FISCHER 344; TISSUE-BRAIN;

RX MEDLINE=96436481; PubMed=8839352;

RA Masters J.N., Cotman S.L., Osterburg H.H., Nichols N.R., Finch C.E.;  
 RT "Modulation of a novel RNA in brain neurons by glucocorticoid and  
 RT mineralocorticoid receptors.";  
 RL Neuroendocrinology 63:28-38(1996).  
 DR EMBL: U25281; AAA87791.1; -;  
 DR InterPro: IPR003124; WH2.  
 DR Pfam: PF02205; WH2; 1.  
 DR SMART: SM00246; WH2; 1.

SQ SEQUENCE 451 AA; 45307 MW; 6E2B50D1876E9CEE CRC64;

Query Match 61.5%; Score 8; DB 11; Length 451;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPPA 11  
 |||||  
 DB 192 PPPPLPPA 199

RESULT 11  
 095763 PRELIMINARY; PRT; 453 AA.  
 AC 095763;

DT 01-MAY-1999 (TREMBLrel. 10, Created)

DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE WUGSC:H.DJ0871815.2 PROTEIN (FRAGMENT).

OS *Homo sapiens* (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99063792; PubMed=9847074;

RA Sulston J.E., Waterston R.;

RT "Toward a complete human genome sequence.";  
 RL Genome Res. 8:1097-1108(1998).  
 RN [2]

RP SEQUENCE FROM N.A.

RA Bourne S., Woliam C., Pieler T.;

RT "The sequence of *Homo sapiens* PAC clone DJ0871815.";  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RA Waterston R.H.;

RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RA Waterston R.H.;

RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.

RN [5]

RP SEQUENCE FROM N.A.

RA Waterston R.;

RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AC004912; AAD15418.1; -;  
 DR InterPro: IPR002965; P\_Rich\_extensn.

DR InterPro: IPR003124; WH2.

DR InterPro: IPR003882; Pistil\_extensin.

DR Pfam: PF02205; WH2; 1.

DR PRINTS; PRO1217; PRICEXTENSIN.  
 DR PRINTS; PRO1218; PSTLEXTENSIN.  
 DR SMART; SM00246; WH2; 1.  
 FT NON\_TER  
 SQ SEQUENCE 453 AA; 46578 MW; 43E3A9C753F97D02 CRC64;

Query Match 61.5%; Score 8; DB 4; Length 453;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 APPPLPP 10  
 Db 261 APPPLPP 268

RESULT 12  
 ID Q920G8 PRELIMINARY; PRT; 485 AA.  
 AC Q920G8;  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE SH3 DOMAIN BINDING PROTEIN.  
 GN CRI6.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 ON NCBI\_TaxID=10116;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FISCHER 344; TISSUE=LIVER;  
 RA Ashman W.H., Smith J.L., Coleman S.L., Masters J.N.;  
 RT "Gene organization and differential splicing of CRI6, a novel gene  
 expressed in CNS neurons.";  
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U31168; AAC99859.1; -;  
 DR EMBL; U31161; AAC99859.1; JOINED.  
 DR EMBL; U31162; AAC99859.1; JOINED.  
 DR EMBL; U31163; AAC99859.1; JOINED.  
 DR EMBL; U31164; AAC99859.1; JOINED.  
 DR EMBL; U31165; AAC99859.1; JOINED.  
 DR EMBL; U31166; AAC99859.1; JOINED.  
 DR EMBL; U31167; AAC99859.1; JOINED.  
 DR EMBL; U31159; AAC99858.1; -;  
 DR InterPro; IPR002965; P\_rich\_extensin.  
 DR InterPro; IPR003124; WH2.  
 DR InterPro; IPR003882; Pfam1\_extensin.  
 DR Pfam; PF02205; WH2; 1.  
 DR PRINTS; PRO1217; PRICEXTENSIN.  
 DR PRINTS; PRO1218; PSTLEXTENSIN.  
 DR SMART; SM00246; WH2; 1.  
 SQ SEQUENCE 485 AA; 49380 MW; 195673B854CB37D9 CRC64;

Query Match 61.5%; Score 8; DB 11; Length 485;  
 Best Local Similarity 100.0%; Pred. No. 1.2;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLPP 11  
 Db 192 PPPPLPP 199

RESULT 13  
 ID Q9FU62 PRELIMINARY; PRT; 507 AA.  
 AC Q9FU62;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE P013PF10.12 PROTEIN.  
 GN P013PF10.12.

OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 ON NCBI\_TaxID=4530;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
 clone: P013PF10.";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP002523; BAB17066.1; -;  
 SQ SEQUENCE 507 AA; 53021 MW; 1B550AFB74164442 CRC64;

Query Match 61.5%; Score 8; DB 10; Length 507;  
 Best Local Similarity 100.0%; Pred. No. 1.2;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 APPPLPP 10  
 Db 404 APPPLPP 411

RESULT 14  
 ID Q9SN05 PRELIMINARY; PRT; 564 AA.  
 AC Q9SN05;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE HYPOTHETICAL 66.6 KDA PROTEIN.  
 GN F3A4.210.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 ON NCBI\_TaxID=3702;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RA Barques M., Collado M.C., Navarro P., Terol J., Perez-Alonso M.,  
 RA Mewes H.W., Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL132978; CAB62121.1; -;  
 DR InterPro; IPR002965; P\_rich\_extensin.  
 DR PRINTS; PRO1217; PRICEXTENSIN.  
 KW Hypothetical protein.  
 SQ SEQUENCE 564 AA; 66602 MW; AD6578D575E1A2BB CRC64;

Query Match 61.5%; Score 8; DB 10; Length 564;  
 Best Local Similarity 100.0%; Pred. No. 1.4;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLPP 11  
 Db 51 PPPPLPP 58

RESULT 15  
 ID Q9W1X3 PRELIMINARY; PRT; 569 AA.  
 AC Q9W1X3;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE CG9848 PROTEIN.  
 GN CG9848.

OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OC NCBI\_TaxId=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Abmayyan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Bens P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Maltel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svrtk R., Tector C., Turner R., Ventler E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003459; AAG22198.1;  
 DR FlyBase: FBgn0034799; CG9848.  
 SQ SEQUENCE 569 AA; 63909 MW; D1D7043B8337A5C CRC64;

Query Match 61.5%; Score 8; DB 5; Length 569;  
 Best Local Similarity 100.0%; Pred. No. 1.4;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 3 APPPLPP 10  
 |||||  
 Db 501 APPPLPP 508

Search completed: January 22, 2002, 16:36:45  
 Job time: 345 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2002, 16:30:53 ; Search time 22.77 Seconds  
(without alignments)  
12.848 Million cell updates/sec

Title: US-09-485-529-103

Perfect score: 13

Sequence: 1 LNAPPPPLPPAPQ 13

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 212252 seqs, 22503292 residues

Word size : 0

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued Patents, AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCRTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53.8	15	2	US-08-769-745-31	Sequence 31, Appl
2	53.8	15	4	US-08-602-999A-356	Sequence 356, App
3	53.8	15	4	US-08-602-999A-414	Sequence 414, App
4	53.8	18	4	US-08-602-999A-320	Sequence 320, App
5	53.8	28	2	US-08-459-568-58	Sequence 58, Appl
6	53.8	28	2	US-08-399-411-58	Sequence 58, Appl
7	53.8	28	3	US-08-769-745-39	Sequence 39, Appl
8	53.8	28	3	US-08-516-859A-58	Sequence 58, Appl
9	53.8	69	4	US-08-818-112-78	Sequence 78, Appl
10	53.8	140	3	US-08-850-961-12	Sequence 12, Appl
11	53.8	141	3	US-08-850-961-10	Sequence 10, Appl
12	53.8	281	2	US-08-810-453-2	Sequence 2, Appl
13	53.8	281	3	US-08-815-190A-2	Sequence 2, Appl
14	53.8	281	4	US-09-290-640-25	Sequence 25, Appl
15	53.8	281	4	US-09-479-524-3	Sequence 3, Appl
16	53.8	281	5	PCT-US95-00362-2	Sequence 2, Appl
17	53.8	288	3	US-08-545-196B-21	Sequence 21, Appl
18	53.8	288	4	US-09-028-327-3	Sequence 3, Appl
19	53.8	311	4	US-09-179-558-66	Sequence 66, Appl
20	53.8	445	2	US-08-900-148-2	Sequence 2, Appl
21	53.8	453	1	US-08-417-330A-18	Sequence 18, Appl
22	53.8	539	3	US-08-906-360-1	Sequence 1, Appl
23	53.8	726	4	US-09-126-980-2	Sequence 2, Appl
24	53.8	726	4	US-09-476-482-2	Sequence 2, Appl
25	53.8	834	3	US-08-539-205A-6	Sequence 6, Appl
26	53.8	933	4	US-08-764-870-14	Sequence 14, Appl
27	53.8	933	4	US-08-980-115-14	Sequence 14, Appl

28	7	53.8	1876	2	US-08-609-049A-12	Sequence 12, Appl
29	7	53.8	1876	2	US-08-609-049A-28	Sequence 28, Appl
30	7	53.8	1876	4	US-09-170-996-12	Sequence 12, Appl
31	7	53.8	1876	4	US-09-170-996-28	Sequence 28, Appl
32	7	53.8	1958	1	US-07-945-283-2	Sequence 2, Appl
33	7	53.8	2205	1	US-08-093-453B-2	Sequence 2, Appl
34	6	46.2	7	1	US-08-230-047-40	Sequence 40, Appl
35	6	46.2	7	2	US-08-769-745-26	Sequence 26, Appl
36	6	46.2	7	5	PCT-US94-01840-11	Sequence 11, Appl
37	6	46.2	8	2	US-08-612-857-8	Sequence 8, Appl
38	6	46.2	10	1	US-08-230-047-12	Sequence 12, Appl
39	6	46.2	10	1	US-08-212-190A-8	Sequence 8, Appl
40	6	46.2	10	2	US-08-900-321-8	Sequence 8, Appl
41	6	46.2	10	3	US-08-899-595-8	Sequence 8, Appl
42	6	46.2	10	4	US-09-001-984C-34	Sequence 34, Appl
43	6	46.2	10	5	PCT-US95-03610-8	Sequence 8, Appl
44	6	46.2	11	1	US-08-336-343A-25	Sequence 25, Appl
45	6	46.2	11	4	US-08-652-877-22	Sequence 22, Appl

#### ALIGNMENTS

RESULT 1  
US-08-769-745-31  
; Sequence 31, Application US/08769745  
; Patent No. 5955259  
; GENERAL INFORMATION:  
; APPLICANT: Holmes, Todd C.  
; APPLICANT: Levitan, Irwin B.  
; APPLICANT: Brandeis University  
; TITLE OF INVENTION: Mechanism for the Regulation of Ion  
; FILE REFERENCE: BR096-02  
; CURRENT APPLICATION NUMBER: US/08/769,745  
; CURRENT FILING DATE: 1996-12-19  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 31  
; LENGTH: 15  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-08-769-745-31

Query Match 53.8%; Score 7; DB 2; Length 15;  
Best Local Similarity 100.0%; Pred. No. 0.28;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 4 PPPPLPP 10  
Db 3 PPPPLPP 9  
RESULT 2  
US-08-602-999A-356  
; Sequence 356, Application US/08602999A  
; Patent No. 6184205  
; GENERAL INFORMATION:  
; APPLICANT: SPARKS, Andrew B.  
; APPLICANT: KAY, Brian K.  
; APPLICANT: THORN, Judith M.  
; APPLICANT: QUILIAM, Lawrence A.  
; APPLICANT: DER, Channing J.  
; APPLICANT: FOWLER, Dana M.  
; APPLICANT: RIDER, James E.  
; TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF  
; TITLE OF INVENTION: ISOLATING AND USING SAME  
; NUMBER OF SEQUENCES: 467  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York

STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/602,999A  
FILING DATE: 16-FEB-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-202  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 356:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-602-999A-356

Query Match 53.8%; Score 7; DB 4; Length 15;  
Best Local Similarity 100.0%; Pred. No. 0.28;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLP 10  
| | | | |  
Db 6 PPPPLP 12

RESULT 3  
US-08-602-999A-414  
Sequence 414, Application US/08602999A  
Patent No. 6184205  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
APPLICANT: KAY, Brian K.  
APPLICANT: THORN, Judith M.  
APPLICANT: QUILLIAM, Lawrence A.  
APPLICANT: DER, Channing J.  
APPLICANT: FOWLKES, Dana M.  
APPLICANT: RIDER, James E.  
TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF  
ISOLATING AND USING SAME  
NUMBER OF SEQUENCES: 467  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/602,999A  
FILING DATE: 16-FEB-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-202

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 414:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-602-999A-414

Query Match 53.8%; Score 7; DB 4; Length 15;  
Best Local Similarity 100.0%; Pred. No. 0.28;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLP 10  
| | | | |  
Db 3 PPPPLP 9

RESULT 4  
US-08-602-999A-320  
Sequence 320, Application US/08602999A  
Patent No. 6184205  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
APPLICANT: KAY, Brian K.  
APPLICANT: THORN, Judith M.  
APPLICANT: QUILLIAM, Lawrence A.  
APPLICANT: DER, Channing J.  
APPLICANT: FOWLKES, Dana M.  
APPLICANT: RIDER, James E.  
TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF  
ISOLATING AND USING SAME  
NUMBER OF SEQUENCES: 467  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/602,999A  
FILING DATE: 16-FEB-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-202  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 320:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-602-999A-320

Query Match 53.8%; Score 7; DB 4; Length 18;  
Best Local Similarity 100.0%; Pred. No. 0.33;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY 4 PPPPLP 10  
| | | | |  
DB 9 PPPPLP 15

## RESULT 5

US-08-459-568-58  
; Sequence 58, Application US/08459568  
; Patent No. 5811304  
; GENERAL INFORMATION:  
; APPLICANT: Huang, Shi  
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting  
; TITLE OF INVENTION: Zinc Finger Proteins  
; NUMBER OF SEQUENCES: 93  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/459,568  
; FILING DATE: 02-JUN-1995  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/399,411  
; FILING DATE: 06-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 58:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 28 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; US-08-459-568-58

Query Match 53.8%; Score 7; DB 2; Length 28;  
Best Local Similarity 100.0%; Pred. No. 0.5;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLP 10  
| | | | |  
DB 15 PPPPLP 21

## RESULT 6

US-08-399-411-58  
; Sequence 58, Application US/08399411  
; Patent No. 5831008  
; GENERAL INFORMATION:  
; APPLICANT: Huang, Shi  
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting  
; TITLE OF INVENTION: Zinc Finger Proteins  
; NUMBER OF SEQUENCES: 93  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA

ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/399,411  
; FILING DATE: 06-MAR-1995  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LJ 1264  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 58:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 28 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; US-08-399-411-58

Query Match 53.8%; Score 7; DB 2; Length 28;  
Best Local Similarity 100.0%; Pred. No. 0.5;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLP 10  
| | | | |  
DB 15 PPPPLP 21

## RESULT 7

US-08-769-745-39  
; Sequence 39, Application US/08769745  
; Patent No. 5955259  
; GENERAL INFORMATION:  
; APPLICANT: Holmes, Todd C.  
; APPLICANT: Levitan, Irwin B.  
; APPLICANT: Brandeis University  
; TITLE OF INVENTION: Mechanism for the Regulation of Ion  
; FILE REFERENCE: BR096-02  
; CURRENT APPLICATION NUMBER: US/08/769,745  
; CURRENT FILING DATE: 1996-12-19  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 39  
; LENGTH: 28  
; TYPE: PRT  
; ORGANISM: Rat  
; US-08-769-745-39

Query Match 53.8%; Score 7; DB 2; Length 28;  
Best Local Similarity 100.0%; Pred. No. 0.5;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLP 10  
| | | | |  
DB 10 PPPPLP 16

## RESULT 8

US-08-516-859A-58  
; Sequence 58, Application US/08516859A  
; Patent No. 6069231  
; GENERAL INFORMATION:  
; APPLICANT: Huang, Shi  
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting  
; TITLE OF INVENTION: Zinc Finger Proteins

NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell & Flores LLP  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/516,859A  
FILING DATE: 18-AUG-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/399,411  
FILING DATE: 06-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/292,683  
FILING DATE: 18-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-LJ 1776  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 58:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 28 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-516-859A-58

Query Match 53.8%; Score 7; DB 3; Length 28;  
Best Local Similarity 100.0%; Pred. No. 0.5;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 PPPPLPP 10  
|||  
Db 15 PPPPLPP 21

RESULT 9  
US-08-818-112-78  
Sequence 78, Application US/08818112  
Patent No. 6290969  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Dillon, Davin C.  
APPLICANT: Campos-Neto, Antonio  
APPLICANT: Houghton, Raymond  
APPLICANT: Vedvick, Thomas S.  
APPLICANT: Twardzik, Daniel R.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS  
NUMBER OF SEQUENCES: 153  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US//08/818,112  
FILING DATE: 13-MAR-1997  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Makl, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.411C6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 78:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 69 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-818-112-78

Query Match 53.8%; Score 7; DB 4; Length 69;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 PPLPPAP 12  
|||  
Db 23 PPLPPAP 29

RESULT 10  
US-08-850-961-12  
Sequence 12, Application US/08850961  
Patent No. 6013517  
GENERAL INFORMATION:  
APPLICANT: Respass, James G.  
APPLICANT: De Polo, Nicholas J.  
APPLICANT: Chada, Sunil  
APPLICANT: Sauter, Sybille  
APPLICANT: Bodner, Mordechai  
APPLICANT: Driver, David A.  
TITLE OF INVENTION: CROSSLESS RETROVIRAL VECTORS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Chiron Corporation, Intellectual Property - R440  
STREET: P.O. Box 8097  
CITY: Emeryville  
STATE: California  
COUNTRY: USA  
ZIP: 94662-8097  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/850,961  
FILING DATE: 05-MAY-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Kruse, No. 6013517man J.  
REGISTRATION NUMBER: 35,235  
REFERENCE/DOCKET NUMBER: 930049.424C4 / 1147.005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 601-3520  
TELEFAX: (510) 655-3542  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 140 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-850-961-12

Query Match 53.8%; Score 7; DB 3; Length 140;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLPP 10  
1111111  
DB 105 PPPPLPP 111

## RESULT 11

US-08-850-961-10  
; Sequence 10, Application US/08850961  
; Patent No. 6013517  
; GENERAL INFORMATION:  
; APPLICANT: Respass, James G.  
; APPLICANT: De Polo, Nicholas J.  
; APPLICANT: Chada, Sunil  
; APPLICANT: Sauter, Sybille  
; APPLICANT: Bodner, Mordchai  
; APPLICANT: Driver, David A.  
; TITLE OF INVENTION: CROSSLESS RETROVIRAL VECTORS  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation, Intellectual Property - R440  
; STREET: P.O. Box 8097  
; CITY: Emeryville  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94662-8097  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/850,961  
; FILING DATE: 05-MAY-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kruse, No. 6013517man J.  
; REGISTRATION NUMBER: 35,235  
; REFERENCE/DOCKET NUMBER: 930049.424C4 / 1147.005  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 601-3520  
; TELEFAX: (510) 655-3542  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 141 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-850-961-10

Query Match 53.8%; Score 7; DB 3; Length 141;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLPP 10  
1111111  
DB 105 PPPPLPP 111

## RESULT 12

US-08-810-453-2  
; Sequence 2, Application US/08810453  
; Patent No. 5858990  
; GENERAL INFORMATION:  
; APPLICANT: Walslh, Kenneth  
; TITLE OF INVENTION: FAS LIGAND COMPOSITIONS FOR TREATMENT OF  
; PROLIFERATIVE DISORDERS

NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
STREET: Federal Reserve Plaza, 600 Atlantic Avenue  
CITY: Boston  
STATE: MA

COUNTRY: US  
ZIP: 02210-2211

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/810,453

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Plumer, Elizabeth R.

REGISTRATION NUMBER: 36,637

REFERENCE/DOCKET NUMBER: S1237/7004

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)720-3500

TELEFAX: (617)720-2441

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 281 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-810-453-2

Query Match 53.8%; Score 7; DB 2; Length 281;  
Best Local Similarity 100.0%; Pred. No. 4.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PPPPLPP 10  
1111111  
DB 50 PPPPLPP 56

## RESULT 13

US-08-815-190A-2  
; Sequence 2, Application US/08815190A  
; Patent No. 6046310  
; GENERAL INFORMATION:  
; APPLICANT: Queen, Cary L.  
; APPLICANT: Schneider, William P.  
; APPLICANT: Vasquez, Maximiliano  
; TITLE OF INVENTION: Fas Ligand Fusion Proteins and their  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/815,190A  
; FILING DATE: 11-MAR-1997  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/614,584  
; FILING DATE: 13-MAR-1996

ATTORNEY/AGENT INFORMATION:  
NAME: Apple, Randolph T.  
REGISTRATION NUMBER: 36,429  
REFERENCE/DOCKET NUMBER: 011823-006710US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 281 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-815-190A-2

Query Match 53.8%; Score 7; DB 3; Length 281;  
Best Local Similarity 100.0%; Pred. No. 4.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPP 10  
|||||  
DB 50 PPPPLPP 56

RESULT 14  
US-09-290-640-25  
Sequence 25, Application US/09290640  
Patent No. 6204055  
GENERAL INFORMATION:  
APPLICANT: Dean, Nicholas M.  
TITLE OF INVENTION: Antisense Compound Modulation of Fas Mediated Signaling  
FILE REFERENCE: ISPH-0351  
CURRENT APPLICATION NUMBER: US/09/290,640  
CURRENT FILING DATE: 1999-04-12  
NUMBER OF SEQ ID NOS: 85  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 25  
LENGTH: 281  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-290-640-25

Query Match 53.8%; Score 7; DB 4; Length 281;  
Best Local Similarity 100.0%; Pred. No. 4.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPP 10  
|||||  
DB 50 PPPPLPP 56

RESULT 15  
US-09-479-524-3  
Sequence 3, Application US/09479524  
Patent No. 6268350  
GENERAL INFORMATION:  
APPLICANT: Barbera-Guillem, Emilio  
TITLE OF INVENTION: Polynucleotides For Inhibiting Metastasis And Tumor Cell Growth  
FILE REFERENCE: B-37  
CURRENT APPLICATION NUMBER: US/09/479,524  
CURRENT FILING DATE: 2000-01-07  
EARLIER APPLICATION NUMBER: US 09/170,948  
EARLIER FILING DATE: 1998-10-13  
EARLIER APPLICATION NUMBER: US 60/062,733  
EARLIER FILING DATE: 1997-10-23  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: WordPerfect  
SEQ ID NO 3  
LENGTH: 281  
TYPE: PRT

ORGANISM: Homo sapiens  
US-09-479-524-3

Query Match 53.8%; Score 7; DB 4; Length 281;  
Best Local Similarity 100.0%; Pred. No. 4.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 PPPPLPP 10  
|||||  
DB 50 PPPPLPP 56

Search completed: January 22, 2002, 16:30:54  
Job time: 274 sec



